

## COMPARISON THEOREMS AND ORBIT COUNTING IN HYPERBOLIC GEOMETRY

MARK POLLICOTT AND RICHARD SHARP

ABSTRACT. In this article we address an interesting problem in hyperbolic geometry. This is the problem of comparing different quantities associated to the fundamental group of a hyperbolic manifold (e.g. word length, displacement in the universal cover, etc.) asymptotically. Our method involves a mixture of ideas from both “thermodynamic” ergodic theory and the automaton associated to strongly Markov groups.

### 0. INTRODUCTION

In [5] Cannon showed that for fundamental groups of many manifolds of negative curvature (including, for example, compact manifolds) the generating function for word length is a rational function. To explain the implications of this property, we recall that the growth of the quantity  $N(n) = \text{Card}\{g \in \Gamma : |g| = n\}$  was studied by Milnor in his fundamental paper [13], where he obtained estimates using comparison with the growth of volume in the universal cover. Cannon used purely combinatorial methods to show that the generating function is rational. In particular, this implies that we can find constants  $\beta_i$  and  $C_i$ , and positive integers  $k_i$  ( $i = 1, \dots, N$ ) such that  $N(n) = \sum_{i=1}^N C_i n^{k_i} \beta_i^n$ .

The key step in Cannon’s approach was to associate to the group an automaton. In this article we shall augment this with the notion of “weighting”. This allows us to draw upon the well-known theory of *Thermodynamic Formalism* to prove a number of new results which can be viewed as weighted analogues of the above results. It has long been understood that this approach must have close connections with symbolic dynamics, Markov partitions, subshifts of finite type (cf. [1], [20], [21] and, in particular, the work of Bourdon [4]) and by extension to the whole paraphernalia of the area of dynamical systems collectively called “Thermodynamic Formalism”. In this article, we shall elaborate on this connection as an integral part of our analysis.

Let  $\mathbb{H}^n$ , for some  $n \geq 2$ , denote  $n$ -dimensional hyperbolic space (i.e. the unique  $n$ -dimensional simply connected, complete Riemannian manifold with all sectional

---

Received by the editors May 23, 1995.

1991 *Mathematics Subject Classification*. Primary 20F32, 22E40, 58E40; Secondary 11F72, 20F10, 58F20.

*Key words and phrases*. Strongly Markov, hyperbolic group, Kleinian group, orbit counting function, negative curvature, Poincaré series.

The first author was supported by The Royal Society through a University Research Fellowship. The second author was supported by the UK SERC under grant number GR/G51930 held at Queen Mary and Westfield College.

curvatures equal to  $-1$ ), and let  $\Gamma \subset Isom(\mathbb{H}^n)$  be a subgroup of the group of isometries of  $\mathbb{H}^n$ . To each element  $g \in \Gamma$  we can associate two values:

1. The *displacement*  $d(x, gx)$ , where  $d$  denotes the Poincaré metric on  $\mathbb{H}^n$  and  $x \in \mathbb{H}^n$  is an arbitrarily chosen point;
2. The *word length*  $|g|$ , relative to a fixed set of generators (cf. section 1)

A fundamental insight due to Milnor was that for co-compact groups, i.e. groups for which the quotient  $\mathbb{H}^n/\Gamma$  is compact, the ratio of word length and displacement is uniformly bounded from above and below (i.e. there exist constants  $A, B > 0$  such that  $A \leq \frac{|g|}{d(x, gx)} \leq B$ , for all  $g \in \Gamma$ ). Our main result is the following more refined comparison theorem.

**Theorem.** *Let  $\Gamma$  be a co-compact group of isometries of  $\mathbb{H}^n$  then for any  $x \in \mathbb{H}^n$ , the following “average” exists*

$$\mathcal{A} = \lim_{T \rightarrow +\infty} \frac{\sum_{\{g \in \Gamma : d(x, gx) \leq T\}} \frac{|g|}{d(x, gx)}}{\sum_{\{g \in \Gamma : d(x, gx) \leq T\}} 1}.$$

Moreover, there exists a constant  $C > 0$  such that

$$\sum_{\{g \in \Gamma : d(x, gx) \leq T\}} \frac{|g|}{d(x, gx)} \sim Ce^{(n-1)T}$$

as  $T \rightarrow +\infty$ .

*Notational Comment.* Given two functions  $A(T)$  and  $B(T)$  we write  $A(T) \sim B(T)$  as  $T \rightarrow +\infty$  if  $\lim_{T \rightarrow +\infty} \frac{A(T)}{B(T)} = 1$ .

This theorem follows from Theorem 1 in section 4.

We briefly summarize the contents of this article. In section 1, we give the basic theory of strongly Markov groups and automata and describe their relationship to hyperbolic geometry. In section 2 we describe an optional hypothesis on automata which holds in many important special cases and allows us to give both simpler proofs and stronger results. (Unfortunately, since we do not know that this property holds for the automata associated to all groups in which we are interested, we have to develop a more general approach in the subsequent sections.) In section 3, we show that the displacement function satisfies an appropriate Hölder condition. In section 4, we state our principal result comparing other weightings (such as word length) to the displacement weighting.

In sections 5-9 we develop the tools needed to prove Theorem 1. In the fifth section we formulate the link between negatively curved groups and the thermodynamic theory by associating to the strongly Markov structure a subshift of finite type. In section 6, we develop the theory of *transfer operators* for these subshifts of finite type. (This involves the extra complication of working with non-mixing subshifts of finite type, in the absence of the extra hypothesis of section 2). In section 7 and section 8 we recall the classical definition of Poincaré series associated to groups of isometries and extend this to introduce certain weighted Poincaré series. We go on to develop their analysis through transfer operators for the subshifts of finite type associated to the automata of hyperbolic groups. In particular, in section 7, we prove a new result on their meromorphic domain. In section 9, we complete the proofs of the principal results stated in section 4.

In the last section we turn to other applications of our basic approach. In section 10, we consider the analogous problem of comparing other weightings with that by word length (under the additional hypothesis of section 4).

The authors would like to thank D. Epstein, J. Parker, S. J. Patterson, I. Redfern and C. Series for valuable conversations.

## 1. WEIGHTED STRONGLY MARKOV GROUPS

Let  $\Gamma$  be an (infinite) finitely presented group with identity element  $e \in \Gamma$ . Let  $\Gamma_0 \subset \Gamma$  be a finite symmetric set of generators for  $\Gamma$  (i.e.  $\Gamma_0$  generates  $\Gamma$  and if  $g \in \Gamma_0$  then  $g^{-1} \in \Gamma_0$ ).

We can define the *word length* of an element  $g \in \Gamma$  to be the *least* number of elements from  $\Gamma_0$  whose concatenation equals  $g$ . We denote this value by  $|g|$ .

**Definition.** We call  $\Gamma$  *strongly Markov* if for any finite symmetric set of generators  $\Gamma_0$  there exists a finite directed graph  $G$  with:

- (i) a vertex set  $V$ ;
- (ii) an edge set  $E \subset V \times V$ ;
- (iii) a distinguished vertex  $*$  in  $V$  such that no edge in  $E$  ends at  $*$ ;
- (iv) a labelling of the edges  $\lambda : E \rightarrow \Gamma_0$ ,

such that there is a bijection between:

- (a) finite paths  $\gamma$  in the graph starting with the distinguished state  $*$ ; and
- (b) elements  $g \in \Gamma$

which associates to the path  $\gamma$  along concurrent edges  $(*, v_1), (v_1, v_2), \dots, (v_{n-1}, v_n)$  the group element  $g = \lambda(*, v_1)\lambda(v_1, v_2) \dots \lambda(v_{n-1}, v_n)$  given by group multiplication of the labelling of the associated edges. Moreover, the word length of  $g$  exactly equals the number of edges in the path (i.e.  $|g| = n$ ).

We shall refer to the graph  $G$  with this additional structure as an automaton.

*Remark.* This definition corresponds to that of “fortement Markov” in the book of Ghys and de la Harpe [8]. We should observe that the assumption  $E \subset V \times V$  is no additional restriction since for any graph where vertices are connected by more than one edge we can associate a new graph which retains the important properties required in the definition of strongly Markov, and in addition has only one edge relating each vertex. More precisely, for the new graph we define the new vertex set to be the edge set of the original graph and then we allow a pair of (the new) vertices to be linked by a (new) edge if in the original graph the corresponding edges were linked by a vertex (with the correct orientations). We can label the edges of the new graph by looking at the vertex they enter, and using the labeling of this vertex (viewed as an edge in the original graph). This is a very familiar device in coding theory of subshifts of finite type.

The most important examples of strongly Markov groups arise in hyperbolic geometry, i.e. certain groups of isometries of hyperbolic space. The limit set  $\Lambda_\Gamma$  of a subgroup  $\Gamma \subset \text{Isom}(\mathbb{H}^n)$  is the set of accumulation points of the set  $\{gx : g \in \Gamma\}$  (for any  $x \in \mathbb{H}^n$ ) in the boundary  $S^{n-1}$ . We are only interested in non-elementary groups i.e. those for which the limit set is infinite. We say that a (non-elementary) group  $\Gamma$  is *convex co-compact* if the quotient of the convex hull of  $\Lambda_\Gamma$  by the action of  $\Gamma$  is compact. Any convex co-compact group is a strongly Markov group. This result is due to Cannon [5] (at least in the co-compact case); a particularly clear account of the proof is given in [8].

We wish to apply these ideas to understand not just the abstract group, but also some of the original geometry. To this end, we want to supplement the basic theory of strongly Markov groups  $\Gamma$  by considering an additional object, namely a weighting, as we define below.

Given an infinite discrete group  $\Gamma$  we shall be interested in *weight functions* i.e. maps  $w : \Gamma \rightarrow \mathbb{R}^+$ . We shall concern ourselves exclusively with weightings that satisfy the following natural comparison condition (and in the sequel whenever we refer to a weighting, we mean one with this property).

*Standing assumption.* For some choice of symmetric generating set  $\Gamma_0 \subset \Gamma$ , there exists a constant  $c > 1$  such that the weighting  $w : \Gamma \rightarrow \mathbb{R}^+$  satisfies  $\frac{1}{c}|g| \leq w(g) \leq c|g|$  for all  $g \in \Gamma$ .

The two most important weightings for groups of isometries are:

- (i) *word length*, usually denoted  $|g|$  which is the length of the shortest presentation of  $g \in \Gamma$  in terms of elements of  $\Gamma_0$ , which only depends on the group properties (and the choice of generators);
- (ii) *displacement*, defined for any point  $x \in \mathbb{H}^n$  by  $w(g) = d(x, gx)$ , which is a natural “geometrical” weighting. The fact that displacement satisfies the standing assumption for co-compact groups is just a restatement of Milnor’s comparison result [13]. For more general convex co-compact groups the result holds unchanged.

## 2. THE STRUCTURE OF AUTOMATA

In this section we want to describe an open problem, the positive solution to which would simplify our analysis. Indeed, the desired property *does* hold in a number of important cases. However, in subsequent sections we shall employ a more complicated analysis, in the absence of a general solution.

**Definition.** Given a directed graph  $G$  we define an equivalence relation  $\sim$  on the vertex set  $V$  by  $i \sim j$  if and only if there is a path in the graph from  $i$  to  $j$  and a path from  $j$  to  $i$ . (If two vertices are in the same equivalence class, then we say that they *communicate*).

An equivalence class with exactly one element is called a *singleton*. For example, the equivalence class  $[*]$  is a singleton.

There is one particular additional assumption on this equivalence relation  $\sim$  which would prove particularly useful for us in subsequent sections.

**Definition.** We shall call an automaton *gregarious* when:

1. all but one of the equivalence classes is a singleton;
2. the largest equivalence class has the property that there exists  $N > 0$  such that for every two vertices  $i$  and  $j$  in this equivalence class there exists a path of length  $N$  from  $i$  to  $j$  and a path of length  $N$  from  $j$  to  $i$ .

We shall call the pair  $(\Gamma, \Gamma_0)$ , where  $\Gamma$  is a group with a preferred symmetric choice of generators  $\Gamma_0$ , *gregarious* if there is a gregarious automaton representing them.

*Question.* For which strongly Markov groups  $\Gamma$  can we find a symmetric set of generators  $\Gamma_0$  such that the pair  $(\Gamma, \Gamma_0)$  is gregarious ?

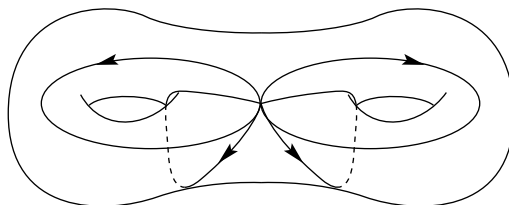


FIGURE 1(A)

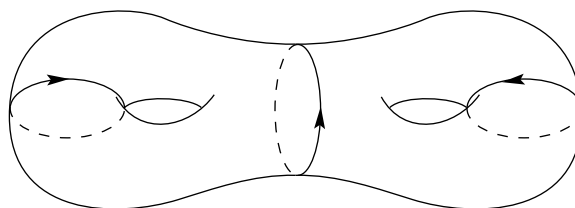


FIGURE 1(B)

We recall that a Fuchsian group is a discrete subgroup of the group of isometries of two dimensional hyperbolic space, and that a Kleinian group is a discrete subgroup of the group of isometries of three dimensional hyperbolic space. A Schottky group is either a Fuchsian or Kleinian group which (as an abstract group) is isomorphic to a free group. For more details cf. [3].

These examples have a bearing on the above question.

**Examples.** When  $\Gamma$  is a Schottky group or a convex co-compact Fuchsian group (with the symmetric set of generators described below) then there are automata which have precisely two equivalence classes, the first corresponding to a singleton for the identity, the second corresponding to all of the rest. For Schottky groups it is easy to see this since it is a straightforward exercise to write down an automaton for a free group with the standard generators [8]. For convex co-compact Fuchsian the coding corresponding to the description in [20], [21] and [1] suffices.

In the case of co-compact Fuchsian groups  $\Gamma$  (without torsion) the Series-Adler-Flatto approach to coding leads to two natural choices of symmetric generators  $\Gamma_0 \subset \Gamma$ . The first choice corresponds to the generators in the standard presentation for the group (i.e.  $a_1 b_1 a_1^{-1} b_1^{-1} \dots a_k b_k a_k^{-1} b_k^{-1} = 1$ ) and to the homotopy classes of the curves illustrated in Figure 1(A).

The second choice of  $\Gamma_0$  corresponds to the side identifications of the fundamental region arising in the “pair of pants” decomposition of the surface  $\mathbb{H}^2/\Gamma$ , as illustrated in Figure 1(B).

### 3. DISPLACEMENT AS A HÖLDER WEIGHTING

In this section we want to show that the *displacement* function on a convex co-compact group of isometries of  $\mathbb{H}^n$  is a Hölder weight function, in the sense described below.

**Definition.** For each element  $a \in \Gamma$  we can define the difference operator by

$$(D_a w)(g) = w(g) - w(ag).$$

In general, this difference may not be particularly small.

**Definitions.** We let  $(\cdot, \cdot)$  denote the *Gromov product* in  $\Gamma$  (with respect to  $\Gamma_0$ ) defined by  $(g, h) := \frac{1}{2}(|g| + |h| - |g^{-1}h|)$  [9].

We say that a weighting  $w : \Gamma \rightarrow \mathbb{R}$  satisfies the *Hölder weight condition* if  $\exists C > 0, 0 < \theta < 1$ , such that  $\forall g, h \in \Gamma$  and  $\forall a \in \Gamma_0$  we have  $|(D_a w)(g) - (D_a w)(h)| \leq C\theta^{(g,h)}$ .

**Example.** In the simple case where  $\Gamma$  is a free group and the elements of  $\Gamma_0$  are the natural generators and their inverses, the Gromov product of two elements  $a_1 \dots a_n, b_1 \dots b_m \in \Gamma$  (written in reduced form) is simply the index of the first term at which their presentation in terms of the generators (and inverses) begin to disagree.

A more geometrical interpretation of the Gromov product is given by noting that  $\Gamma$  there exist constants  $C_1, C_2 > 0$  such that for  $g \neq h$  we have

$$C_1 \leq \frac{(g, h)}{d(x, [gx, hx])} \leq C_2,$$

where  $x \in \mathbb{H}^n$  is a given point in hyperbolic space, and  $[gx, hx]$  is the geodesic segment linking  $gx$  and  $hx$  [6].

*Remark.* For the Hölder weight condition to be natural, we need to know that it is independent of the choice of the set  $\Gamma_0$ . This is true because if  $\Gamma_1$  is another choice of generators then there exists constants  $C_1, C_2, K > 0$  such that

$$C_1(g, h)_{\Gamma_0} - K \leq (g, h)_{\Gamma_1} \leq C_2(g, h)_{\Gamma_0} + K$$

for every  $g, h \in \Gamma$ .

**Proposition 1.** *Consider a convex co-compact group  $\Gamma$  of isometries acting on  $\mathbb{H}^n$ , and consider a symmetric set of generators  $\Gamma_0$ . The weighting  $w : \Gamma \rightarrow \mathbb{R}$  defined by  $w(g) = d(x, gx)$ , for  $g \in \Gamma$ , is Hölder.*

*Proof.* We work in the Poincaré ball model and without loss of generality assume that  $x = 0$ . We begin by a simple estimate on the distance  $d(0, g0)$  (in the hyperbolic metric) from 0 to  $g0$ . We recall that the Poincaré metric (along radial lines) is given by

$$\begin{aligned} ds^2 &= \frac{4(dx_1^2 + \dots + dx_n^2)}{(1 - x_1^2 - \dots - x_n^2)^2} \\ &= \frac{4dr^2}{(1 - r^2)^2} = \frac{4dr^2}{(1 + r)^2(1 - r)^2} \end{aligned}$$

where  $r^2 = x_1^2 + \dots + x_n^2$  and observe that on the Poincaré ball we have that  $\frac{1}{2} \leq \frac{1}{1+r} \leq 1$ .

We can write

$$\begin{aligned} (3.1) \quad D_a(d(0, g0)) &= d(0, g0) - d(0, ag0) \\ &= 2 \int_0^{|g0|} \frac{1}{1 - r^2} dr - 2 \int_0^{|ag0|} \frac{1}{1 - r^2} dr \\ &= 2 \int_{|ag0|}^{|g0|} \frac{1}{1 - r^2} dr \\ &= \log \left( \frac{1 - |ag0|}{1 - |g0|} \frac{1 + |g0|}{1 + |ag0|} \right) \end{aligned}$$

Also we observe that  $\frac{1}{1+|g0|} = \frac{1}{2} + E(g)$ , where  $E(g) = \frac{1-|g0|}{2(1+|g0|)} = O(1-|g0|)$ . Thus we have the equation

$$\begin{aligned} & D_a(d(0, g0)) - D_a(d(0, h0)) \\ &= \log \left( \frac{1-|ag0|}{1-|g0|} \frac{1-|h0|}{1-|ah0|} \right) + \log \left( \frac{1+|g0|}{1+|ag0|} \frac{1+|ah0|}{1+|h0|} \right) \\ &= \log \left( \frac{1-|ag0|}{1-|g0|} \frac{1-|h0|}{1-|ah0|} \right) \\ &\quad - \log \left( \frac{1}{2} + E(ag) \right) + \log \left( \frac{1}{2} + E(g) \right) \\ &\quad - \log \left( \frac{1}{2} + E(h) \right) + \log \left( \frac{1}{2} + E(ah) \right) \\ &= \log \left( \frac{1-|ag0|}{1-|g0|} \frac{1-|h0|}{1-|ah0|} \right) + O(1-|g0|, 1-|h0|). \end{aligned}$$

The function  $a$  has an analytic extension to the closed  $n$ -ball (and thus, in particular, is  $C^2$ ), and so we can write

$$\frac{1-|ag0|}{1-|g0|} = \det(a')(x) + O(1-|g0|)$$

where  $x \in S^{n-1}$  is a point on the boundary which satisfies  $|x - ag0| = O(1-|g0|)$ . Similarly, we can estimate  $\frac{1-|ah0|}{1-|h0|} = \det(a')(x) + O(1-|h0|)$  where  $x \in S^{n-1}$  is any point on the boundary which satisfies  $|x - ah0| = O(1-|h0|)$ .

Consider the images  $g0, h0 \in \mathbb{H}^n$ . We recall that  $C_1 \leq \frac{(g,h)}{d(0,[g0,h0])} \leq C_2$  for constants  $C_1, C_2 > 0$ , say. Moreover, we have the simple bounds for any  $z \in \mathbb{H}^n$  that

$$-\log(1-|z|) \leq d(0, z) = 2 \int_0^{|z|} \frac{dr}{1-r^2} \leq -2\log(1-|z|).$$

Let  $C$  be the unique geodesic in the disc which passes through the two points  $g0$  and  $h0$  (and therefore  $[g0, h0] \subset C$ ). In particular,  $C$  is a euclidean circular arc which meets the unit circle perpendicularly. Let  $z \in C$  be the point on the geodesic arc  $C$  which is closest to 0 (i.e.  $d(0, z) = d(0, C)$ ).

We have two cases to consider. In the first case we assume that  $z \in [g0, h0]$  then

$$-2\log(1-|z|) \geq d(0, z) = d(0, [g0, h0]) \geq \frac{1}{C_2}(g, h)$$

and we see that  $(1-|z|) \leq e^{-\frac{1}{2C_2}(g,h)}$ . In particular, we may take  $x$  to be any point on the boundary interior to the arc  $C$ .

In the second case we assume  $z \notin [g0, h0]$  then we see that  $d(0, g0) = d(0, [g0, h0])$  by monotonicity of the distance  $c \rightarrow d(0, c)$  of points on  $c \in C$  from 0. If we now take the circular arc  $C'$  which passes through  $g0$  and for which  $g0$  is the nearest point to 0, then we see that it contains  $h0$  in its interior. In particular, we see that

$$-2\log(1-|g0|) \geq d(0, g0) = d(0, [g0, h0])$$

and thus  $(1-|g0|) \leq e^{-\frac{1}{2C_2}(g,h)}$ . In particular, we can take  $x$  to be any point on the unit circle interior to the arc  $C'$ .

$$\begin{aligned}
& D_a(d(0, g0)) - D_a(d(0, h0)) \\
&= \log \left( \frac{1 - |ag0|}{1 - |g0|} \frac{1 - |h0|}{1 - |ah0|} \right) + O((1 - |g0|), (1 - |h0|)) \\
&= \log \left( \frac{\det(a')(x)}{\det(a')(x)} \right) + O((1 - |g0|), (1 - |h0|)) \\
&= O((1 - |g0|), (1 - |h0|)) \\
&\leq C\theta^{(g,h)}
\end{aligned}$$

(where  $\theta = e^{-\frac{1}{2C_2}}$ , say).

This completes the proof.  $\square$

#### 4. COMPARISON THEOREMS FOR WEIGHTINGS

In this section we shall state a more general form of the theorem given in the introduction. In the sections 5-9 we shall concern ourselves with the proof.

Our main theorem describes an asymptotic average of a Hölder weighting relative to the displacement  $d(x, gx)$ .

**Theorem 1** (Asymptotic average for displacement). *Let  $\Gamma$  be a convex co-compact non-elementary group of isometries of  $\mathbb{H}^n$  with  $\delta = HD(\Lambda_\Gamma) > 0$  denoting the Hausdorff dimension of the limit set  $\Lambda_\Gamma$  of  $\Gamma$ . Let  $\Gamma_0$  be a finite symmetric set of generators for  $\Gamma$ . Given a Hölder weighting  $w : \Gamma \rightarrow \mathbb{R}$  there exists a constant  $C_x > 0$  such that*

$$\sum_{\{g \in \Gamma : d(x, gx) \leq T\}} \frac{w(g)}{d(x, gx)} \sim C_x e^{\delta T}$$

as  $T \rightarrow +\infty$ . Equivalently, the following limiting average exists

$$\mathcal{A} = \lim_{T \rightarrow +\infty} \frac{\sum_{\{g \in \Gamma : d(x, gx) \leq T\}} \frac{w(g)}{d(x, gx)}}{\sum_{\{g \in \Gamma : d(x, gx) \leq T\}} 1}.$$

Notice that with the particular choice  $w(g) = |g|$  (and restricting to co-compact groups), this gives the Main Theorem.

We now present a second result that gives an asymptotic comparison for any Hölder weighting relative to word length.

**Proposition 2** (Asymptotics for word length). *Let  $\Gamma$  be a strongly Markov group with exponential growth. Given a (Hölder) weighting  $w : \Gamma \rightarrow \mathbb{R}$ , there exists constants  $\beta_i$  (of equal modulus),  $C_i$  and positive integers  $k_i$  ( $i = 1, \dots, N$ ) such that*

$$\sum_{\{g \in \Gamma : |g| = n\}} \frac{w(g)}{|g|} \sim \sum_{i=1}^N C_i n^{k_i} \beta_i^n.$$

*Remark.* There is a slight asymmetry in our statements, since we are not claiming any asymptotic average where we order group elements by word length. The reason is that we do not know that the asymptotics of the counting function correspond to the asymptotics of the comparison function. However, for *gregarious* pairs  $(\Gamma, \Gamma_0)$  we can make this assertion. We shall return to this point later.



## 5. STRONGLY MARKOV GROUPS AND ASSOCIATED SUB-SHIFTS OF FINITE TYPE

We now begin with some technical results which will provide an important link between the theory of strongly Markov groups and the thermodynamic theory of subshifts of finite type.

Given the vertex set  $V$  for the directed graph  $G$  introduced in section 1 we want to add an extra state 0 to form a larger set  $V \cup \{0\}$ . We form a new graph by adding to  $E$  an edge from  $i$  to 0 for each  $i \in V - \{*\}$  and an edge from 0 to itself. We adopt the convention that  $\lambda(i, 0) = e$  (the identity element in  $\Gamma_0$ )  $\forall i \in V \cup \{0\} - \{*\}$ . (See Figure 2.) We associate a square matrix  $A$  with zero-one entries whose columns and rows are indexed by  $V \cup \{0\}$  and whose entries are determined by:

1.  $A(i, j) = 1$  if  $(i, j) \in E$ ;
2.  $A(i, j) = 1$  if  $i \in (V \cup \{0\}) - \{*\}$  and  $j = 0$ ;
3.  $A(*, *) = 0$ ;
4.  $A(i, j) = 0$  in all the remaining cases.

Let

$$X_A = \{x = (x_n) \in \prod_{n \geq 0} \{1, \dots, d\} : A(x_n, x_{n+1}) = 1 \quad \forall n \geq 0\}$$

and let  $\sigma : X_A \rightarrow X_A$  be the shift defined by  $(\sigma x)_n = x_{n+1}$ .

We define a metric on  $X_A$  by

$$d(x, y) = \sum_{n=0}^{+\infty} \frac{1 - \delta_{x_n, y_n}}{2^{|n|}}$$

where  $\delta_{ij}$  is the standard Kronecker delta.

We now see how to embed  $\Gamma$  into the subshift  $X_A$ . We associated to each group element  $g \in \Gamma$  the *unique* path  $\gamma$  which passes through concurrent vertices  $*, a_1, a_2, \dots, a_n \in V$  with  $n = |g|$  and such that  $\lambda(*, a_1)\lambda(a_1, a_2) \dots \lambda(a_{n-1}, a_n) = g$ . We then define an inclusion  $\iota : \Gamma \rightarrow X_A$  by

$$\iota(g) = (*, a_1, a_2, \dots, a_n, 0, 0, \dots).$$

If we define  $X_A^0 \subset X_A$  to be those sequences which end with an infinite string of zeros, then it is easy to see that  $X_A^0$  is dense in  $X_A$ . This is because an open set contains a cylinder set, i.e. a set of the form

$$C = \{x \in X_A : x_i = y_i, \forall 0 \leq i \leq n-1\}$$

for some admissible string  $y_0, \dots, y_{n-1}$ , which in turn contains a sequence which ends in an infinite string of zeros.

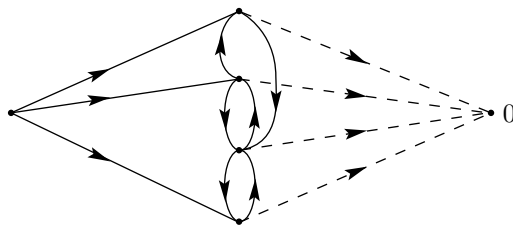


FIGURE 2

**Lemma 1.** *A weighting  $w : \Gamma \rightarrow \mathbb{R}$  satisfying the Hölder weight condition defines a Hölder continuous function  $f : X_A^0 \rightarrow \mathbb{R}$  by*

$$f(x_0, 0, \dots) = -w(e),$$

$$f(x_0, x_1, x_2, \dots, x_n, 0, 0, \dots) = (D_{\lambda(x_0, x_1)^{-1}w}) (\lambda(x_0, x_1)\lambda(x_1, x_2) \dots \lambda(x_{n-1}, x_n)).$$

*This then extends uniquely to a Hölder continuous function  $f : X_A \rightarrow \mathbb{R}$ .*

*Proof.* It suffices to consider sequences

$$x = (x_0, x_1, \dots, x_{n-1}, 0, 0, \dots),$$

$$y = (y_0, y_1, \dots, y_{m-1}, 0, 0, \dots)$$

where  $n, m \geq 2$ .

We assume that  $x, y$  are close in as much as  $x_i = y_i$  for  $i = 1, \dots, k-1 \leq \min\{n, m\}$  but  $x_k \neq y_k$ .

Observe that

$$\begin{aligned} |f(x) - f(y)| &= |D_{\lambda(x_0, x_1)^{-1}w}(\lambda(x_0, x_1) \dots \lambda(x_{n-2}, x_{n-1})) \\ &\quad - D_{\lambda(y_0, y_1)^{-1}w}(\lambda(y_0, y_1) \dots \lambda(y_{m-2}, y_{m-1}))| \\ &\leq C\theta^{(\lambda(x_0, x_1) \dots \lambda(x_{n-2}, x_{n-1}), \lambda(y_0, y_1) \dots \lambda(y_{n-2}, y_{n-1}))} \\ &\leq C\theta^{\frac{1}{2}((n-1)+(m-1)-((n-1)+(m-1)-2k))} \\ &= C\theta^k. \end{aligned}$$

This shows that  $f : X_A^0 \rightarrow \mathbb{R}$  is Hölder continuous. Since  $X_A^0 \subset X_A$  is dense, it follows from uniform continuity that  $f$  uniquely extends to a Hölder continuous function on  $X_A$ . This completes the proof of the lemma.  $\square$

By our definition of  $f$  we have the identity  $f^n(\iota(g)) = w(g)$ , where  $n = |g|$ .

In Proposition 1 we showed that displacement gave a Hölder weighting for convex co-compact subgroups  $\Gamma$  of the isometries of  $n$ -dimensional hyperbolic space. In this case, we shall denote by  $r : X_A \rightarrow \mathbb{R}$  the function given by Lemma 1, i.e.

$$\begin{aligned} r(x_0, 0, \dots) &= 0, \\ (5.1) \quad r(x_0, x_1, x_2, \dots, x_n, 0, 0, \dots) &= d(0, \lambda(x_0, x_1)\lambda(x_1, x_2) \dots \lambda(x_{n-1}, x_n)0) \\ &\quad - d(0, \lambda(x_1, x_2) \dots \lambda(x_{n-1}, x_n)0). \end{aligned}$$

## 6. STRONGLY MARKOV GROUPS AND TRANSFER OPERATORS.

In this section we shall introduce transfer operators associated to the subshifts introduced in the previous section. The main technical difficulty that we must overcome is that the shifts may not be topologically mixing (i.e. the matrices  $A$  may not be aperiodic) or even topologically transitive.

A non-negative square matrix  $M$  is called *irreducible* if for each pair of indices  $i, j$  there exists a positive integer  $n$  such that  $M^n(i, j) > 0$ . Such a matrix  $M$  is called *aperiodic* if exists a positive integer  $n$  such that  $M^n(i, j) > 0$  for each pair of indices  $i, j$ .

A subshift of finite type  $\sigma : X_A \rightarrow X_A$  is called (*topologically*) *transitive* if there exists a dense orbit. Furthermore, we have the stronger property of (*topological*) *mixing* if for each pair of non-empty open sets  $U, V \subset X_A$  there exists  $n > 0$  such

that  $\sigma^n U \cap V \neq \emptyset$ . The shift  $\sigma : X_A \rightarrow X_A$  is transitive if and only if  $A$  is irreducible and it is mixing if and only if  $A$  is aperiodic.

Consider an arbitrary  $d \times d$  matrix  $A$  with zero-one entries. Following [19] we can choose the indexing of the rows and columns so that  $A$  takes the form

$$A = \begin{pmatrix} B_{11} & 0 & \cdots & 0 \\ B_{21} & B_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ B_{k1} & B_{k2} & \cdots & B_{kk} \end{pmatrix}$$

where  $B_{ij}$  for  $i \geq j$  are simply sub-matrices which are irreducible. Let  $X_A$  be the associated subshift of finite type.

We define a new matrix by

$$B = \begin{pmatrix} B_{11} & 0 & \cdots & 0 \\ 0 & B_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B_{kk} \end{pmatrix}$$

and let  $X_B \subset X_A$  be the associated subshift of finite type.

We shall use the symbol  $\sigma$  to denote the shift on either of the spaces.

We denote by  $C^\alpha(X_A, \mathbb{C})$  the Banach space of  $\alpha$ -Hölder continuous functions with the usual norm  $\|f\|_\alpha = |f|_\infty + |f|_\alpha$ , where  $|f|_\infty$  is the usual supremum norm and

$$|f|_\alpha = \sup \left\{ \frac{|f(x) - f(y)|}{d(x, y)^\alpha} : x, y \in X_A, x \neq y \right\}.$$

Similarly, we denote by  $C^\alpha(X_B, \mathbb{C})$  the Banach space of  $\alpha$ -Hölder continuous functions on the space  $X_B$ . The inclusion  $X_B \subset X_A$  gives rise to a projection  $\pi : C^\alpha(X_A) \rightarrow C^\alpha(X_B)$ . Given  $f \in C^\alpha(X_A)$  we can denote the image  $\pi(f) = f'$ .

**Definition.** On each space of Hölder continuous functions we can define *Ruelle transfer operators*  $L_f : C^\alpha(X_A) \rightarrow C^\alpha(X_A)$  by

$$(L_f h)(x) = \sum_{\substack{\sigma y = x \\ y \neq \dot{0}}} e^{f(y)} h(y),$$

where  $\dot{0} = (0, 0, 0, \dots)$ , and  $L_{f'} : C^\alpha(X_B) \rightarrow C^\alpha(X_B)$  by

$$(L_{f'} h)(x) = \sum_{\substack{\sigma y = x \\ y \neq \dot{0}}} e^{f'(y)} h(y).$$

*Remark.* This definition of the Ruelle transfer operator differs from the usual one in that in the summation over pre-images  $y$  of  $x$ , we exclude the possibility  $y = \dot{0}$ . However, it agrees with the more familiar definition for all  $x \neq \dot{0}$  and its only effect on the spectrum is to exclude an eigenvalue  $e^{f(\dot{0})}$  (corresponding to the eigenvector which is the characteristic function for the set  $\{\dot{0}\}$ ). This change is necessary in order for us to be able to write the Poincaré series  $\eta(s)$ , defined in section 7, as a sum of iterates of transfer operators.

For a bounded linear operator  $T : B \rightarrow B$  acting on a Banach space  $B$ , we define the *essential spectrum*  $\text{ess}(T)$  to be the subset of the spectrum  $\text{spec}(T) \subset \mathbb{C}$  of  $T$  consisting of those  $\lambda \in \text{spec}(T)$  such that at least one of the following is true:

1.  $\text{Range}(\lambda - T)$  is not closed in  $B$ ;
2.  $\lambda$  is a limit point of  $\text{spec}(T)$ ;
3.  $\bigcup_{r=1}^{\infty} \ker(\lambda - T)^r$  is infinite dimensional.

We define the *essential spectral radius* to be  $\rho_e(T) = \sup\{|\lambda| : \lambda \in \text{ess}(T)\}$ . The operator  $T : B \rightarrow B$  is *quasi-compact* if the essential spectral radius is strictly smaller than the spectral radius.

**Lemma 2.** 1. *The operators  $L_f : C^\alpha(X_A) \rightarrow C^\alpha(X_A)$  and  $L_{f'} : C^\alpha(X_B) \rightarrow C^\alpha(X_B)$  are both quasi-compact.*  
 2. *The isolated eigenvalues of both  $L_f : C^\alpha(X_A) \rightarrow C^\alpha(X_A)$  and  $L_{f'} : C^\alpha(X_B) \rightarrow C^\alpha(X_B)$  coincide.*

*Proof.* The proof of this lemma depends on some technical estimates. Part (1) is familiar in the case where  $A$  and  $B$  are irreducible or aperiodic. The proof in that case applies here too. The key point is the use of the following Basic Identities

$$\|L_f^n h\|_\alpha \leq \left( \text{Const.} |h|_\infty + \left(\frac{1}{2}\right)^{n\alpha} |h|_\alpha \right) \|L_{Ref}^n 1\|_\alpha$$

and

$$\|L_{f'}^n h\|_\alpha \leq \left( \text{Const.} |h|_\infty + \left(\frac{1}{2}\right)^{n\alpha} |h|_\alpha \right) \|L_{Ref'}^n 1\|_\alpha$$

(cf. [18, p. 89], [15], ch. 2). Notice that in [15] it is assumed that the matrices are aperiodic and it is also assumed for simplicity that the operators satisfy the normalization condition that  $L_{Ref} 1 = 1$ . However, it is easy to see that this is not essential provided the inequality is amended to the form above). The proof of these identities is given in [15] and although there we are working with the hypothesis that the shift was mixing, it is apparent from the simple proof that this is not essential. The proof of quasi-compactness is then as follows. If we let  $h_n = E_n(h)$  be an approximation to  $h$  on cylinders of length  $n$  then we have that  $|h - h_n|_\infty \leq \|h\|_\alpha \left(\frac{1}{2}\right)^{n\alpha}$  and  $|h - h_n|_\alpha \leq \|h\|_\alpha$ . This shows that

$$\begin{aligned} \|(L_f^n - L_f^n \circ E_n)h\|_\alpha &= \|L_f^n(h - h_n)\|_\alpha \\ &\leq \left( \text{Const.} |h - h_n|_\alpha + \left(\frac{1}{2}\right)^{n\alpha} |h - h_n|_\alpha \right) \|L_{Ref}^n 1\|_\alpha \\ &\leq \left( \text{Const} \left(\frac{1}{2}\right)^{n\alpha} \|h\|_\alpha + \|h\|_\alpha \left(\frac{1}{2}\right)^{n\alpha} \right) \|L_{Ref}^n 1\|_\alpha. \end{aligned}$$

It then follows from a result of Nussbaum that the essential spectral radius of the operator  $L_f$  is smaller than  $\left(\frac{1}{2}\right)^\alpha \rho(L_{Ref})$  where  $\rho(L_{Ref})$  is the spectral radius of  $L_f$  [14]. Precisely the same reasoning applies to  $L_{f'}$ .

For part (2) we can consider the operators

$$L_f, L_{f_n} : C^\alpha(X_A) \rightarrow C^\alpha(X_A)$$

and

$$L_{f'}, L_{f'_n} : C^\alpha(X_B) \rightarrow C^\alpha(X_B),$$

where  $f_n = E_n(f)$ .

We can increase  $\alpha$  to  $\alpha' > \alpha$  so that the essential spectral radius changes by only a small amount [10].

It is well-known that

$$\|L_f - L_{f_n}\|_{\alpha'} \rightarrow 0 \text{ and } \|L_{f'} - L_{f'_n}\|_{\alpha'} \rightarrow 0$$

as  $n \rightarrow +\infty$  [18, pp. 91, 94]. Moreover, since the isolated eigenvalues for the operator acting on  $C^\alpha$  could only accumulate at the essential spectral radius, the isolated eigenvalues in  $C^{\alpha'}$  are contained in the isolated eigenvalues in  $C^\alpha$ , and the essential spectral radius on  $C^{\alpha'}$  is strictly bigger than on  $C^\alpha$ , we deduce that the spectrum of  $L_f : C^{\alpha'} \rightarrow C^{\alpha'}$  has only a finite number of isolated eigenvalues. In particular, using perturbation theory, it follows that the spectra of  $L_f, L_{f_n} : C^\alpha(X_A) \rightarrow C^\alpha(X_A)$  can be made arbitrarily close (and similarly for  $L_{f'}, L_{f'_n} : C^\alpha(X_B) \rightarrow C^\alpha(X_B)$ ).

We can recode the each of the subshifts of finite type such that the new states are simply words  $\tilde{x} = x_0 x_1 \dots x_{n-2}$  of length  $n-1$ . Given the locally constant functions  $f_n(x_0, \dots, x_{n-1})$  and  $f'_n(x_0, \dots, x_{n-1})$  we can now write these as functions of two co-ordinates in the re-coded shifts i.e.  $f_n(x) = f_n(\tilde{x}, \tilde{y})$  and  $\pi f_n(x) = \pi f_n(\tilde{x}, \tilde{y})$ . Moreover, we can introduce matrices  $M_n(\tilde{x}, \tilde{y}) = e^{f_n(\tilde{x}, \tilde{y})}$  and  $P_n(\tilde{x}, \tilde{y}) = e^{f'_n(\tilde{x}, \tilde{y})}$ , with the convention that the entry is taken to be zero if the transition from the symbol  $\tilde{x}$  to the symbol  $\tilde{y}$  is not allowed.

The effect of the recoding on  $B$  is that the transition matrix on words of length  $n-1$  still consists of blocks down the diagonal. In particular, we can choose the indexing of the rows and columns so as to have the presentation

$$P_n = \begin{pmatrix} P_{11} & 0 & \dots & 0 \\ 0 & P_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & P_{l_k l_k} \end{pmatrix}$$

for some  $l_k$ . Moreover, we can also assume that the matrix  $M_n$  takes the form

$$M_n = \begin{pmatrix} P_{11} & 0 & \dots & 0 \\ M_{21} & P_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ M_{l_k 1} & M_{l_k 2} & \dots & P_{l_k l_k} \end{pmatrix}.$$

The isolated eigenvalues in the spectra of the operators  $L_{f_n}$  and  $L_{f'_n}$  (acting on  $C^{\alpha'}(X_A)$  and  $C^{\alpha'}(X_B)$ ) are precisely the eigenvalues of the associated matrices  $M_n$  and  $P_n$  [18].

We first show that the eigenvalues of the matrices  $A_n$  and  $B_n$  must coincide. Since the eigenvalues can be determined from the asymptotics of the values

$$\begin{aligned} \text{trace}(P_n^m) &= \text{trace}(P_{11}^m) + \dots + \text{trace}(P_{l_k l_k}^m) \\ &= \text{trace}(M_n^m) \end{aligned}$$

as  $m \rightarrow +\infty$  we immediately see that the eigenvalues of the matrices  $P_n$  and  $M_n$  are the same.

By our observations above, we see that the associated operators  $L_{f_n} : C^{\alpha'}(X_A) \rightarrow C^{\alpha'}(X_A)$  and  $L_{f'_n} : C^{\alpha'}(X_B) \rightarrow C^{\alpha'}(X_B)$  have the same *isolated* eigenvalues.

For the transfer operators  $L_f : C^{\alpha'}(X_A) \rightarrow C^{\alpha'}(X_A)$  and  $L_{f'} : C^{\alpha'}(X_B) \rightarrow C^{\alpha'}(X_B)$  we can use the approximation in norm by the operators  $L_{f_n}$  and  $L_{f'_n}$  and perturbation theory to conclude the equivalence of the isolated eigenvalues for the original operators.

This completes the proof of the lemma.  $\square$

Lemma 2 shows that the spectral properties of the transfer operator are “carried” by the restriction to the transitive components in the subshift. Let  $X_i$  for  $i = 1, \dots, k$  be the transitive components of the shift (corresponding to the diagonal blocks  $A_{ii}$  in the decomposition of  $A$ ).

As is well-known, for each such  $1 \leq i \leq k$  there exists a further finite decomposition  $X_{i1} \cup \dots \cup X_{iN_i}$  with the property that induced shift  $\sigma_i : X_i \rightarrow X_i$  has the effect that

- (i)  $\sigma_i : X_{i,j} \rightarrow X_{i,j+1}$  where  $j$  is taken modulo  $n$ .
- (ii)  $\sigma_i^{N_i} : X_{i,j} \rightarrow X_{i,j}$  is mixing.

An understanding of the spectrum of the Ruelle transfer operator will play a crucial role in our subsequent analysis.

**Definition.** We can associate to each continuous function  $g : X_A \rightarrow \mathbb{R}$  the *pressure*  $P(g) \in \mathbb{R}$  defined by

$$P(g) = \limsup_{n \rightarrow +\infty} \frac{1}{n} \log \left( \sum_{\sigma^n x = x} e^{g(x) + g(\sigma x) + \dots + g(\sigma^{n-1} x)} \right).$$

The pressure is also given by the equivalent variational identity

$$P(g) = \sup \{ h(\mu) + \int g d\mu : \mu \text{ is an invariant probability} \}$$

and as the logarithm of the spectral radius of the operator  $L_g$ . The equivalence of these definitions is well-known for the case of transitive subshifts of finite type. It is easy to see that they are equivalent at our present level of generality by considering the transitive components.

It is easy to see that for  $g > 0$  real valued we have that the map given by  $t \rightarrow \rho(L_{tg})$  from  $\mathbb{R}$  to  $\mathbb{R}$  which associates to  $t$  the spectral radius of the operator  $L_{tg}$  is strictly increasing.

The spectral results for the transfer operator are usually presented in the case of mixing shifts as follows.

**Proposition 3** (Complex Ruelle operator theorem). *Let  $\sigma : X_A \rightarrow X_A$  be a mixing subshift of finite type. Then*

1. *When  $g$  is real valued the operator  $L_g$  has a simple (isolated) maximal positive eigenvalue  $e^{P(g)}$  with an associated strictly positive eigenfunction  $\psi \in C^\alpha(X_A, \mathbb{R})$ . There is a unique probability measure  $\nu$  on  $X_A$  such that  $L_g^* \nu = e^{P(g)} \nu$ . (We can choose the normalization by requiring that  $\int \psi d\nu = 1$ .) Furthermore, the rest of the spectrum is contained in a disk of radius strictly smaller than  $e^{P(g)}$ .*
2. *When  $g$  is complex, the spectral radius of  $L_g$  is less than or equal to  $e^{P(\text{Reg})}$ .*
3. *The spectral radius of  $L_g$  is strictly less than  $e^{P(\text{Reg})}$  unless  $\text{Im}(g) = u \circ \sigma - u + \Psi + a$  where  $u \in C^0(X_A, \mathbb{R})$  and  $\Psi \in C^0(X_A, 2\pi\mathbb{Z})$  and  $a$  is a constant. Furthermore, if such an identity does hold then  $L_g$  has a simple maximal eigenvalue  $e^{P(\text{Reg}) + ia}$  and the rest of the spectrum is contained in a disk of radius strictly less than  $e^{P(\text{Reg})}$ .*

[15, Theorems 2.2, 4.5, and 10.2]

The pressure  $P(g)$  was defined above in the case of real valued functions  $g$ . In part (1) of Proposition 3 we saw that, for mixing systems, it also characterized the simple isolated eigenvalue  $e^{P(g)}$  for the Ruelle operator  $L_g$ . We shall now see how to use perturbation theory to extend this definition to complex valued functions in a neighbourhood of the real valued functions. We begin by quoting a technical lemma.

**Lemma 3** [10, VII.3]. *Let  $B(V)$  denote the Banach algebra of bounded linear operators on a (complex) Banach space  $V$ . If  $L_0 \in B(V)$  has a simple isolated eigenvalue  $\lambda_0$  with corresponding eigenvector  $v_0$  then for any  $\epsilon > 0$  there exists  $\delta > 0$  such that if  $L \in B(V)$  with  $\|L - L_0\| < \delta$  then  $L$  has a simple isolated eigenvalue  $\lambda(L)$  and the corresponding eigenvector  $v(L)$  with  $\lambda(L_0) = \lambda_0$ ,  $v(L_0) = v_0$  and such that*

- (i)  $L \rightarrow \lambda(L)$ ,  $L \rightarrow v(L)$  are analytic for  $\|L - L_0\| < \delta$ , and
- (ii) for  $\|L - L_0\| < \delta$ , we have that  $|\lambda(L) - \lambda_0| < \epsilon$  and  $\text{spec}(L) - \{\lambda(L)\} \subset \{z \in \mathbb{C} : |z - \lambda_0| > \epsilon\}$ .

Moreover, if  $\Sigma_0 = \text{spec}(L_0) - \{\lambda_0\}$  is contained in the interior of a circle  $C$  centred at  $0 \in \mathbb{C}$  then provided  $\delta > 0$  is sufficiently small,  $\Sigma = \text{spec}(L) - \{\lambda(L)\}$  will also be contained in the interior of  $C$ .

If  $g \in C^\alpha(X_A, \mathbb{C})$  is real valued then by part (1) of Proposition 3,  $L_g$  and  $e^{P(g)}$  satisfy the hypothesis of Lemma 3. Combining this with the assertion that the map  $g \rightarrow L_g$  is analytic we see that for  $f$  in a neighbourhood of  $g$  in  $C^\alpha(X_A, \mathbb{C})$  we can define  $e^{P(f)}$  as the perturbed simple isolated eigenvalue of  $L_f$  and that the map  $f \rightarrow e^{P(f)}$  is analytic. We then define the pressure of  $f$  to be the principal branch of  $\log(e^{P(f)})$ .

*Remark.* 1. It is useful to note that (for mixing systems) there is an explicit formula for the first derivative of pressure. Specifically,  $\frac{dP(tg)}{dt}|_{t=a} = \int g d\mu$ , where  $\mu$  is the unique equilibrium state for the Hölder function  $ag$  i.e. the unique probability measure for which the supremum defining the pressure  $P(ag)$  is realized.

2. When the subshift  $\sigma : X_A \rightarrow X_A$  is not mixing then there is still an isolated maximal eigenvalue, however, it may no longer be a simple eigenvalue and the eigenfunctions are not necessarily positive. The main technical difficulty is to relate the general case to that for a mixing shift space.

By considerations in the last section, we understand that the existence of unity as an eigenvalue for the operator  $L_{-s_0 r} : C^\alpha(X_A) \rightarrow C^\alpha(X_A)$  implies that unity is also an eigenvalue for  $L_{-s_0 r'} : C^\alpha(X_B) \rightarrow C^\alpha(X_B)$  corresponding to the transitive components in the shift. If we write  $X_B = X_1 \cup \dots \cup X_k$ , where  $\sigma_i = \sigma|_{X_i}$  is transitive for  $i = 1, \dots, k$  then we can denote the restriction of  $r'$  to  $X_i$  by  $r_i$ . We can identify

$$C^\alpha(X_B) = C^\alpha(X_1) \times \dots \times C^\alpha(X_k)$$

and then write

$$L_{-s_0 r'}(h_1, \dots, h_k) = (L_{-s_0 r_1} h_1, \dots, L_{-s_0 r_k} h_k).$$

This decomposition shows that if  $L_{-s_0 r'}$  has an eigenvalue equal to unity then so must one (or more) of the operators  $L_{-s_0 r_i} : C^\alpha(X_i) \rightarrow C^\alpha(X_i)$ .

Assume that unity is an eigenvalue for the operator  $L_{-s_0 r_i} : C^\alpha(X_i) \rightarrow C^\alpha(X_i)$ . Recall that we can write  $X_i = X_{i1} \cup \dots \cup X_{iN_i}$  and choose  $N_i \geq 1$  such that  $\sigma_i^{N_i}(X_{ij}) \subset X_{ij}$  (for each  $1 \leq j \leq N_i$ ) and, moreover, the map  $\sigma_i^{N_i} : X_{i,j} \rightarrow X_{i,j}$  is mixing.

It follows that unity is *also* an eigenvalue for the  $N_i$ -th iterate  $L_{-s_0 r_i}^{N_i} : C^\alpha(X_i) \rightarrow C^\alpha(X_i)$  (This is, of course, just a trivial consequence of the spectral theorem.) If we introduce the function  $r_i^{N_i}(x) = r(x) + r(\sigma x) + \dots + r(\sigma^{N_i-1}x)$  then we can write  $L_{-s_0 r_i}^{N_i} = \tilde{L}_{-s_0 r_i^{N_i}}$  where now we replace  $\sigma$  by  $\sigma^{N_i}$  in the definition of the Ruelle transfer operator. This is immediate from the definitions.

Finally, since  $\sigma_i^{N_i} : X_{ij} \rightarrow X_{ij}$  is a continuous map we see that under the natural decomposition

$$C^\alpha(X_i) = C^\alpha(X_{i1}) \times \dots \times C^\alpha(X_{iN_i})$$

we can write

$$\tilde{L}_{-s_0 r_i^{N_i}}(h_1, \dots, h_{N_i}) = \left( \tilde{L}_{-s_0 r_i^{N_i}} h_1, \dots, \tilde{L}_{-s_0 r_i^{N_i}} h_{N_i} \right).$$

We thus observe that if unity is an eigenvalue for  $L_{-s_0 r_i^{N_i}} : C^\alpha(X_i) \rightarrow C^\alpha(X_i)$  then it is an eigenvalue on one or more of the components  $L_{-s_0 r_i^{N_i}} : C^\alpha(X_{ij}) \rightarrow C^\alpha(X_{ij})$ .

In the special case of gregarious automata we observe that the associated subshift of finite type  $X_A$  is very close to being mixing. The most important consequence of this is that the spectral theory of the transfer operators  $L_f$  takes exactly the same form as for mixing systems. In particular, the conclusions of Proposition 3 hold.

## 7. MEROMORPHIC EXTENSIONS OF POINCARÉ SERIES

We begin by recalling the classical definition of the Poincaré series associated to a discrete group of isometries  $\Gamma$  of  $\mathbb{H}^n$ .

**Definition.** We define the *Poincaré series* associated to  $\Gamma$  to be the function of a complex variable

$$\eta(s) = \sum_{g \in \Gamma - \{e\}} e^{-sd(x, gx)}$$

for a complex variable  $s \in \mathbb{C}$ , whenever the summation converges. (For convenience, we modify the usual definition to exclude the identity element  $e \in \Gamma$  from the summation.)

If  $\delta = \delta(\Gamma)$  denotes the Hausdorff dimension of  $\Lambda_\Gamma$  then it is a fundamental result that the series converges for  $Re(s) > \delta$  and diverges for  $Re(s) < \delta$  [22] and so defines an analytic function in the half-plane  $Re(s) > \delta$ .

This function is usually studied via the spectral theory of the Laplace-Beltrami operator  $\Delta : L^2(\mathbb{H}^n/\Gamma) \rightarrow L^2(\mathbb{H}^n/\Gamma)$ . Here we employ an alternative approach based on the transfer operators introduced in the preceding section.

To take advantage of our embedding of  $\Gamma$  into the subshift of finite type  $X_A$ , we observe that  $\eta(s)$  may also be written in the following way

$$(7.1) \quad \eta(s) = \sum_{n=1}^{\infty} \sum_{z \in S_n} e^{-sr^n(z\dot{0})}$$

where  $S_n$  denotes the set of all allowed finite paths  $z = z_0 \dots z_n$  of (edge) length  $n$  with  $z_0 = *$  and  $z_i \neq 0$ ,  $i = 1, \dots, n$ , and where  $z\dot{0} = (z_0, \dots, z_n, 0, 0, \dots)$ . The



function  $r : X_A \rightarrow \mathbb{R}$  is defined by (5.1). (We recall that  $r^n(x) = r(x) + r(\sigma x) + \dots + r(\sigma^{n-1}x)$ .)

It is possible to write the Right Hand Side of (7.1) in terms of the transfer operator by means of the next lemma. This will allow us to apply the results of the preceding section. We shall use the following simple identity

**Lemma 4.** *Define  $\chi : X_A \rightarrow \mathbb{R}$  by  $\chi(x) = 1$  if  $x_0 = *$ , and 0 otherwise. For any  $k \in C(X_A)$  we have the following simple identity*

$$\sum_{z \in S_n} e^{k^n(z\dot{0})} = (L_k^n \chi)(\dot{0}), \quad \text{for } n \geq 1.$$

*Proof.* This is by direct computation. For any  $x \in X_A$  and  $k \in C(X_A)$  we can write

$$\sum_{\substack{\sigma^n y = x : y_0 = * \\ y_i \neq 0, i=1, \dots, n}} e^{k^n(y)} = \sum_{\substack{\sigma^n y = x \\ \sigma^{n-1} y \neq \dot{0}}} e^{k^n(y)} \chi(y) = L_k^n \chi(x).$$

The identity follows upon setting  $x = \dot{0}$ . □

Comparing (7.1) and Lemma 4, we can now write

$$(7.2) \quad \eta(s) = \sum_{n=1}^{+\infty} (L_{-sr}^n \chi)(\dot{0})$$

for  $\operatorname{Re}(s) > \delta$ .

From the above, it is clear that  $\delta = \inf\{s \in \mathbb{R} : \rho(L_{-sr}) < 1\}$  and so  $\delta$  is the unique real number such that  $\rho(L_{-\delta r}) = 1$ .

An important consequence of the spectral theory of the transfer operators described in Lemma 2 and Proposition 3 is the following.

**Proposition 4.** *The function  $\eta(s)$  has an extension as a meromorphic function to the half-plane  $\operatorname{Re}(s) > \delta - \epsilon$ , for some  $\epsilon > 0$ .*

*Proof.* By Lemma 2 we have that for any  $\rho' > \rho_e(L_{-sr})$  the operator  $L_{-sr}$  has at most a finite number of isolated eigenvalues of finite multiplicity contained in the annulus  $\rho' \leq |z| \leq \rho(L_{-Re(s)r})$ . Without loss of generality we can assume that the circle  $|z| = \rho'$  is disjoint from the spectrum. For each eigenvalue  $\lambda$  lying in the annulus  $\rho' \leq |z| \leq \rho(L_{-Re(s)r})$  we shall denote by  $\mathbb{P}_\lambda : C^\alpha(X_A, \mathbb{C}) \rightarrow C^\alpha(X_A, \mathbb{C})$  the projection onto the generalized eigenspace associated to  $\lambda$ . Also we shall denote by  $Q : C^\alpha(X_A, \mathbb{C}) \rightarrow C^\alpha(X_A, \mathbb{C})$  the projection associated to the part of the spectrum in  $|z| < \rho'$ . Each of these projections varies analytically with  $s$  and commutes with the operator  $L_{-sr}$ .

We have the following presentation for the iterates of the operator  $L_{-sr}$

$$(7.3) \quad L_{-sr}^n = \sum_{\lambda} \mathbb{P}_\lambda L_{-sr}^n + Q L_{-sr}^n$$

Substituting (7.3) into (7.2) we obtain that

$$(7.4) \quad \begin{aligned} \eta(s) &= \sum_{\lambda} \left( \sum_{n=0}^{\infty} L_{-sr}^n \mathbb{P}_\lambda L_{-sr} \chi \right) (\dot{0}) + \left( \sum_{n=1}^{\infty} L_{-sr}^n Q \chi \right) (\dot{0}) \\ &= \sum_{\lambda} ((I - L_{-sr})^{-1} \mathbb{P}_\lambda L_{-sr} \chi) (\dot{0}) + \left( \sum_{n=1}^{\infty} L_{-sr}^n Q \chi \right) (\dot{0}) \end{aligned}$$

We first observe that the second term in (7.4) converges to an analytic function provided that  $\rho' < 1$ . Since we know by the proof of Lemma 2 that the essential spectral radius  $\rho_e(L_{-sr})$  is bounded above by  $(\frac{1}{2})^\alpha \rho(L_{-Re(s)r})$  and that  $t \rightarrow \rho(L_{-tr})$  is decreasing there exists  $\delta' < \delta$  such that if  $Re(s) > \delta'$  then the condition  $\rho' < 1$  is satisfied.

We next observe that the first term in (7.4) is meromorphic since  $(I - L_{-sr})^{-1}$  can be written in the form  $(I - L_{-sr})^{-1} = \frac{N(s)}{\det(I - L_{-sr})}$  where  $N(s)$  is an analytic operator valued function and, furthermore, it is well known that  $\det(I - L_{-sr})$  is analytic.  $\square$

The domain of meromorphy of  $\eta(s)$  was previously studied via the resolvent kernel of the Laplace-Beltrami operator. In [12] it was shown that this kernel has a meromorphic extension to the entire complex plane. In [16], Patterson used this result to deduce that  $\eta(s)$  has a meromorphic extension to a neighbourhood on the half-plane  $Re(s) \geq \delta$  and went on, using ergodic theoretic ideas, to prove the following Proposition 5. Thus our Proposition 4 can be combined with Patterson's arguments to give a proof of Proposition 5 (and its corollary) which is independent of the approach based on the Laplace-Beltrami operator.

**Proposition 5** [16]. *The function  $\eta(s)$  has an extension as an analytic function to a neighbourhood of  $Re(s) = \delta$  (for some  $\epsilon > 0$ ), apart from a simple pole at  $s = \delta$ . Furthermore, the residue of  $\eta(s)$  at  $s = \delta$  is positive.*

**Corollary 5.1** [16]. *The orbital counting function*

$$N_\Gamma(x, T) = \text{Card}\{g \in \Gamma : d(x, gx) \leq T\}$$

*satisfies the asymptotic formula*

$$N_\Gamma(x, T) \sim Ce^{\delta T}, \quad \text{as } T \rightarrow +\infty,$$

*where  $C$  is a positive constant.*

*Proof.* The corollary is an immediate consequence of Proposition 5 and the Wiener-Ikehara Tauberian theorem.  $\square$

*Remark.* In the case of Fuchsian groups this result was obtained independently by Lalley [11].

## 8. WEIGHTED POINCARÉ SERIES

We shall prove the comparison theorems stated in the introduction and section 4 by means of an analytic method based on the study of certain complex functions. We begin by introducing a “weighted” Poincaré series which generalizes those studied in the last section.

**Definition.** Given a weighting  $w : \Gamma \rightarrow \mathbb{R}^+$  we define (for  $s, z \in \mathbb{C}$ ) the weighted Poincaré series

$$\eta(s, z) = \sum_{g \in \Gamma - \{e\}} e^{-sd(x, gx) + zw(g)}$$

whenever the summation converges. The summation converges provided that  $Re(s) > \delta$  and  $|z|$  is sufficiently small (how small depending on  $s$ ). Note that  $\eta(s, 0) = \eta(s)$ .

Ultimately, this will allow us to study

$$\xi(s) = \frac{\partial}{\partial z} \eta(s, 0) = \sum_{g \in \Gamma - \{e\}} w(g) e^{-sd(x, gx)}$$

which we will use to derive the asymptotics required for Theorem 1. For simplicity we shall use the notation  $\partial_1 = \frac{\partial}{\partial s}$  and  $\partial_2 = \frac{\partial}{\partial z}$ .

If  $f \in C^\alpha(X_A)$  is the function associated to the weighting  $w : \Gamma \rightarrow \mathbb{R}$  (if  $w(g) = |g|$  then clearly  $f = 1$ ), then applying Lemma 4 again gives that

$$(8.1) \quad \eta(s, z) = \sum_{n=1}^{+\infty} (L_{-sr+zf}^n \chi)(\dot{0})$$

for  $\operatorname{Re}(s) > \delta$  and  $|z|$  small.

We now consider the spectrum of  $L_{-sr+zf}$  for  $s$  close to  $\delta$  and  $z$  close to 0. First note that by Proposition 3 that  $L_{-\delta r}$  has 1 as an isolated eigenvalue of finite multiplicity,  $N$  say, with an associated  $N$ -dimensional generalized eigenspace  $V$  of  $C^\alpha(X_A)$ . In fact, as we shall see below, the restriction of  $L_{-\delta r}$  to  $V$  is, in fact, diagonalizable and  $L_{-sr+zf}$  has eigenvalues  $\lambda_1(s, z), \dots, \lambda_N(s, z)$  with  $\lambda_1(\delta, 0) = \dots = \lambda_N(\delta, 0) = 1$ . If we had  $N = 1$ , then by standard perturbation theory for linear operators,  $\lambda_1(s, z)$  would be analytic on a neighbourhood of  $(\delta, 0)$  but if  $N > 1$  then, a priori, all we know is that  $(s, z) \rightarrow \lambda_1(s, z) + \dots + \lambda_N(s, z)$  is analytic [10]. However, by the discussion in the preceding section, each  $\lambda_u(s, z)$ ,  $u = 1, \dots, N$ , occurs as a simple isolated eigenvalue for the operator  $L_{-sr_j+zf_j} : C^\alpha(X_j) \rightarrow C^\alpha(X_j)$  for one of the transitive components  $X_j$  of  $X_B$  and thus we conclude that  $\lambda_1(s, z), \dots, \lambda_N(s, z)$  are all individually analytic on a neighbourhood of  $(\delta, 0)$  (and that their associated one dimensional eigenprojections  $\mathbb{P}_u(s, z)$ ,  $u = 1, \dots, N$ , are also analytic).

In general,  $L_{-\delta r}$  will also have a finite number of other eigenvalues with modulus 1 (in fact, they will all be roots of unity). Although these eigenvalues may not be simple, once again considering transitive components, we see that they vary analytically under perturbation and we shall write  $\lambda_v(s, z)$ ,  $v = N + 1, \dots, M$ , for the perturbation of these eigenvalues.

Finally, provided  $(s, z)$  is sufficiently close to  $(\delta, 0)$  the rest of the spectrum of  $L_{-sr+zf}$  is contained in a disc of radius at most  $1 - \epsilon$ , for some  $\epsilon > 0$ , by Lemma 3. We write  $Q(s, z)$  for the (analytic) projection onto this part of the spectrum.

We may now write iterates of the operator  $L_{-sr+zf}^n$ ,  $n \geq 1$ , in the form

$$(8.2) \quad \begin{aligned} (L_{-sr+zf}^n \chi)(\dot{0}) &= \sum_{u=1}^N \lambda_u(s, z)^{n-1} (\mathbb{P}_u(s, z) L_{-sr+zf} \chi)(\dot{0}) \\ &+ \sum_{v=N+1}^M \lambda_v(s, z)^n (J_{v,n}(s, z) \chi)(\dot{0}) \\ &+ (Q(s, z) L_{-sr+zf}^n \chi)(\dot{0}) \end{aligned}$$

where

1.  $J_{v,n}(s, z) = \sum_{i,j=1}^{N_v} h_i L_{ij}^n \mu_j$ ,
2.  $\{h_i\}_{i=1}^{N_v}$  is a basis of eigenvectors for the generalized eigenspace associated to  $\lambda_v$  and  $\{\mu_i\}_{i=1}^{N_v}$  is a dual basis,

3.  $L = L^{(v)}$  is the  $N_v \times N_v$  Jordan matrix written in the form

$$L = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ \frac{1}{\lambda_v} & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & \frac{1}{\lambda_v} & 1 \end{pmatrix}.$$

Substituting (8.2) into (8.1) we obtain the following expression for  $\eta(s, z)$ ,

$$\begin{aligned} \eta(s, z) &= \sum_u \frac{(\mathbb{P}_u(s, z)L_{-sr+zf}\chi)(\dot{0})}{1 - \lambda_u(s, z)} \\ &\quad + \sum_v \sum_{i,j=1}^{N_v} h_i(\dot{0})\mu_j(\chi) ((I - \lambda_v L)^{-1})_{i,j} + U_1(s, z) \\ (8.3) \quad &= \sum_u \frac{1}{1 - \lambda_u(s, z)} G_u(s, z) \\ &\quad + \sum_v \frac{C_v(s, z)}{(1 - \lambda_v(s, z))^{N_v}} + U_1(s, z) \end{aligned}$$

where  $G_u(s, z) = \mathbb{P}_u(s, z)(L_{-sr+zf}\chi)(\dot{0})$ ,  $C_v(s, z)$  and  $U_1(s, z)$  is analytic for  $(s, z)$  sufficiently close to  $(\delta, 0)$ . (Here we have made use of the well-known fact that the singularities for  $((I - \lambda_v L)^{-1})$  occur at the zeros of  $\det(1 - \lambda_v L) = (1 - \lambda_v(s, z))^{N_v}$ .) Moreover, the summation over  $v$  is also analytic provided no  $\lambda_v(s, z) = 1$  and this cannot occur provided  $(s, z)$  is sufficiently close to  $(\delta, 0)$ . Thus, on a neighbourhood of  $(\delta, 0)$ , we have

$$\eta(s, z) = \sum_u \frac{1}{1 - \lambda_u(s, z)} G_u(s, z) + U_2(s, z)$$

where  $U_2(s, z)$  is analytic.

We can now see why  $L_{-\delta r}$  restricted to  $V$  is diagonalizable. Suppose  $z = 0$  in the above expression, then

$$\eta(s) = \eta(s, 0) = \frac{1}{s - \delta} \sum_u \frac{-1}{\partial_1 \lambda_u(\delta, 0)} G_u(\delta, 0) + U_2(s, 0)$$

agreeing with [16]. One possible complication might be if  $L_{-sr}|_V$  were to have off-diagonal elements in its Jordan Form. However, were this to be the case, then this would give rise to multiple poles for  $\eta(s)$  at  $s = \delta$ . However, we know this not to be the case by Proposition 5.

For our later analysis using the Tauberian theorems, we also need to know about the behaviour of  $\eta(s, z)$  on the whole line  $\operatorname{Re}(s) = \delta$ . Again using Proposition 3, we have a representation for  $L_{-sr+zf}$  as in (8.2). However, setting  $z = 0$ , the result that  $\eta(s, 0)$  has no poles with  $\operatorname{Re}(s) = \delta$ ,  $\operatorname{Im}(s) \neq 0$ , shows that 1 cannot occur as an eigenvalue for  $L_{-sr}$  (and hence not for  $L_{-sr+zf}$  providing  $|z|$  is sufficiently small). Thus we have proved the following proposition.

**Proposition 6.** (i) *The weighted Poincaré series  $\eta(s, z)$  is analytic on a neighbourhood of*

$$(\{s: \operatorname{Re}(s) \geq \delta\} \times \{0\}) - \{(\delta, 0)\}.$$

(ii) In a neighbourhood of  $(\delta, 0)$  we can write

$$\eta(s, z) = \sum_u \frac{1}{1 - \lambda_u(s, z)} G_u(s, z) + U_2(s, z)$$

with  $U_2(s, z)$  analytic.

As we remarked above, we only need to study  $\xi(s) = \partial_2 \eta(s, 0)$ . Clearly, by Proposition 6 (i),  $\xi(s)$  is analytic in a neighbourhood of  $\{s : \operatorname{Re}(s) \geq \delta\} - \{\delta\}$ .

On the other hand, by Proposition 6 (ii),

$$\xi(s) = \sum_u \frac{\partial_2 \lambda_u(s, 0)}{(1 - \lambda_u(s, 0))^2} G_u(s, 0) + \sum_u \frac{1}{1 - \lambda_u(s, 0)} \partial_2 G_u(s, 0) + U_3(s)$$

with  $U_3(s) = \partial_2 U_2(s, 0)$  analytic. Analyticity of the eigenvalues and of  $G_u(s, 0)$  then gives the following result.

**Corollary 6.1.** (i) The function  $\xi(s)$  is analytic in a neighbourhood of  $\{s : \operatorname{Re}(s) \geq \delta\} - \{\delta\}$

(ii) In a neighbourhood of  $s = \delta$ , we have that

$$\begin{aligned} \xi(s) &= \frac{1}{(s - \delta)^2} \sum_u \left( \frac{\partial_2 \lambda_u(\delta, 0)}{(\partial_1 \lambda_u(\delta, 0))^2} \right) G_u(\delta, 0) \\ &\quad + \frac{1}{s - \delta} \sum_u \frac{-\partial_2 G_u(\delta, 0)}{\partial_1 \lambda_u(\delta, 0)} + U_4(s) \end{aligned}$$

with  $U_4(s)$  analytic. Furthermore, the coefficient of  $\frac{1}{(s - \delta)^2}$  is positive.

*Proof.* We only need to show that the coefficient of  $\frac{1}{(s - \delta)^2}$ , namely

$$\sum_u \left( \frac{\partial_2 \lambda_u(\delta, 0)}{(\partial_1 \lambda_u(\delta, 0))^2} \right) G_u(\delta, 0)$$

is strictly greater than zero. This comes from our original hypothesis and comparison with the derivative of the Poincaré series. More specifically, we know from the standing assumption and Milnor's result that there exists  $C > 1$  such that for real valued  $s > \delta$  we can write

$$\begin{aligned} -\frac{1}{C} \frac{d}{ds} \eta(s) &= \frac{1}{C} \sum_{g \in \Gamma} d(x, gx) e^{-sd(x, gx)} \\ &\leq \xi(s) := \sum_{g \in \Gamma} w(g) e^{-sd(x, gx)} \\ &\leq C \sum_{g \in \Gamma} d(x, gx) e^{-sd(x, gx)} \\ &= -C \frac{d}{ds} \eta(s). \end{aligned}$$

Furthermore, we know that the Poincaré series  $\eta(s)$  takes the form  $\eta(s) = \frac{c}{s - \delta} + A(s)$  (for some constant  $c > 0$  and some analytic function  $A(s)$ ) in a neighbourhood of  $s = \delta$ . From this identity we can deduce that  $\frac{d}{ds} \eta(s) = \frac{-c}{(s - \delta)^2} + \frac{d}{ds} A(s)$  and then since both  $\frac{d}{ds} \eta(s)$  and  $\xi(s)$  are meromorphic in a neighbourhood of  $s = \delta$  the above comparison allows us to deduce that  $\xi(s)$  also has a pole at  $s = \delta$  of order two (with coefficient somewhere in the interval  $[\frac{c}{C}, cC]$ ).  $\square$

## 9. PROOFS OF THEOREM 1 AND PROPOSITION 2

We have now introduced all of the ideas required to prove Theorem 1 and Proposition 2.

Before proceeding to the proof of Theorem 1, we need to recall the following result (which appears in a more general form in either [7] Theorem III, or [2]).

**Proposition 7.** *For a monotone increasing function  $\phi : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  we associate the integral*

$$f(s) = \int_0^{+\infty} e^{-sT} d\phi(T)$$

and assume that

- (i)  $f(s)$  is analytic on a neighbourhood of  $\{s \in \mathbb{C} : \operatorname{Re}(s) \geq \delta\} - \{\delta\}$ ,
- (ii) in a neighbourhood of  $s = \delta$  we can write  $f(s) = \frac{r(s)}{(s-\delta)^n} + g(s)$  where  $n > 0$  is a positive integer, and  $g(s), r(s)$  are analytic and non-zero in a neighbourhood of  $s = \delta$ , with  $A = r(\delta) \neq 0$ ,

then

$$\phi(T) \sim \frac{A}{(n-1)!} e^{\delta T} T^{n-1}.$$

*Remark.* When  $n = 1$  this is the more familiar Wiener-Ikehara Tauberian theorem.

*Proof of Theorem 1.* We can write the series  $\xi(s)$  as an integral of the following form

$$\xi(s) = \int_0^\infty e^{-sT} d\pi(T)$$

where we introduce

$$\pi(T) = \sum_{\{g \in \Gamma : d(x, gx) \leq T\}} w(g).$$

For  $\operatorname{Re}(s) > \delta$  we know from Corollary 6.1, that there exists constants  $A, B$  with  $A \neq 0$  such that

$$\xi(s) - \left( \frac{A}{(s-\delta)^2} + \frac{B}{(s-\delta)} \right)$$

is analytic on a neighbourhood of the half-plane  $\operatorname{Re}(s) \geq \delta$ . In particular, in a neighbourhood of  $s = \delta$ , we can write the function in the form  $\xi(s) = \frac{r(s)}{(s-\delta)^2} + g(s)$  where  $g(s)$  is analytic in a neighbourhood of  $s = \delta$ , and  $r(s) = A + B(s-\delta)$ . We first want to use these results on deduce the asymptotics of  $\pi(T)$ . It follows from Proposition 7 that we have the asymptotic formula

$$\pi(T) \sim \frac{A}{6} T e^{\delta T}.$$

We want to use this to deduce the corresponding asymptotic formula for

$$\rho(T) = \sum_{\{g \in \Gamma : d(x, gx) \leq T\}} \frac{w(g)}{d(x, gx)}.$$

We begin by observing that  $\pi(T) \leq T\rho(T)$ . In particular, this allows us to see that

$$\liminf_{T \rightarrow +\infty} \frac{\rho(T)}{\frac{A}{6} e^{\delta T}} \geq \liminf_{T \rightarrow +\infty} \frac{T\rho(T)}{\pi(T)} \geq 1.$$

To get an upper bound, we proceed as follows. We first observe that for any  $0 < \gamma < 1$  we have that

$$(9.1) \quad \begin{aligned} \rho(T) &= \rho(T\gamma) + \sum_{\{T\gamma < d(x, gx) \leq T\}} \frac{w(g)}{d(x, gx)} \\ &\leq \rho(T\gamma) + \frac{1}{\gamma} \frac{\pi(T)}{T}. \end{aligned}$$

This then allows us to write

$$(9.2) \quad \limsup_{T \rightarrow +\infty} \frac{\rho(T)}{\frac{A}{6} e^{\delta T}} \leq \frac{1}{\gamma} \limsup_{T \rightarrow +\infty} \frac{\pi(T)}{\frac{A}{6} e^{\delta T} T} = \frac{1}{\gamma}.$$

Since  $\gamma$  can be chosen arbitrarily close to unity, we deduce from (8.1) and (8.2) that

$$1 \leq \liminf_{T \rightarrow +\infty} \frac{\rho(T)}{\frac{A}{6} e^{\delta T}} \leq \limsup_{T \rightarrow +\infty} \frac{\rho(T)}{\frac{A}{6} e^{\delta T}} \leq 1$$

and so that  $\rho(T) \sim \frac{A}{6} e^{\delta T}$ . To complete the proof of Theorem 1, it suffices to recall from Corollary 5.1 that  $\text{Card}\{g \in \Gamma : d(x, gx) \leq T\} \sim \text{Const.} e^{\delta T}$ . (We refer the reader to [15] for an account of the corresponding situation where one replaces Poincaré series by zeta functions.)  $\square$

In some sense, the proof of Proposition 2 is more elementary, since it does not require Tauberian theorems.

*Proof of Proposition 2.* As we observed in the introduction, Cannon's analysis of the rationality of the generating function for the word length shows that for the function  $N(n) = \text{Card}\{g \in \Gamma : |g| = n\}$  there exist constants  $\beta_i$ ,  $C_i$  and positive integers  $k_i$  ( $i = 1, \dots, N$ ) such that

$$N(n) = \sum_{i=1}^N C_i n^{k_i} \beta_i^n.$$

The values  $\beta_i$  are the eigenvalues of the transition matrix  $A$  and  $k_i$  is determined by the size of the Jordan blocks.

We need estimates on a weighted counting function of the form

$$M(n) = \sum_{y \in S_n} f^n(y\dot{0}) = \sum_{|g|=n} w(g)$$

where  $f \in C^\alpha(X_A)$  is the function associated to the weighting  $w$ . The most convenient approach is to introduce the complex function

$$\begin{aligned} \beta(z, s) &= \sum_{n=1}^{\infty} z^n \sum_{y \in S_n} e^{s f^n(y\dot{0})} \\ &= \sum_{n=1}^{\infty} z^n (L_{sf}^n \chi(\dot{0})) \end{aligned}$$

where we have used Lemma 4. The hypothesis that  $\Gamma$  has exponential growth implies that the transition matrix  $A$  has spectral radius greater than one.

For  $s$  sufficiently small, we know by Lemma 3 that the spectrum of  $L_{sg}$  will remain close to the spectrum of  $A$  in the sense that  $\forall \epsilon > 0$  sufficiently small  $\exists \delta > 0$  such that for  $|s| < \delta$  the operator  $L_{sf}$  has isolated eigenvalues  $\lambda_v(s)$  where

1.  $\lambda_v(0)$  are the eigenvalues of maximum modulus  $\rho$  for  $A$  and they occur with multiplicity  $N_v$ ,
2.  $|\lambda_v(0) - \lambda_v(s)| < \epsilon$ ,  $\forall v$ ,
3. The rest of the spectrum of  $L_{sf}$  is contained in a disc of radius  $\rho - 2\epsilon$ .

Furthermore, although the eigenvalues  $\lambda_v$  are not necessarily simple they occur as simple eigenvalues for transitive components of  $X_A$  and thus are analytic in  $s$ .

By analogy with the spectral presentation of the operator in (8.2), we have a presentation of the form

$$L_{sf}^n \chi(\dot{0}) = \sum_v \lambda_v^n(s) (J_{v,n}(s) \chi)(\dot{0}) + (Q(s) L_{sf}^n \chi)(\dot{0})$$

where

1.  $J_{v,n}(s) = \sum_{i,j=1}^{N_v} h_i L_{ij}^n \mu_j$ ,
2.  $\{h_i\}_{i=1}^{N_v}$  is a basis of eigenvectors for the generalized eigenspace associated to  $\lambda_v$  and  $\{\mu_i\}_{i=1}^{N_v}$  is a dual basis,
3.  $L$  is the  $N_v \times N_v$  Jordan matrix written in the form

$$L = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ \frac{1}{\lambda_v} & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & \frac{1}{\lambda_v} & 1 \end{pmatrix},$$

4.  $Q(s)$  is the projection operator associated to the rest of the spectrum.

In particular, we can write that

$$\begin{aligned} \beta(z, s) &= \sum_{n=0}^{\infty} \sum_v (z^n \lambda_v^n(s) (J_{v,n}(s) \chi)(\dot{0})) \\ &\quad + \sum_{n=1}^{\infty} z^n (Q(s) L_{sf}^n \chi)(\dot{0}) \\ &= \sum_v \sum_{i,j=1}^{N_v} \left( \sum_{n=1}^{\infty} z^n \lambda_v^n(s) (L^n)_{ij} h_i(\dot{0}) \mu_j(\chi) \right) + C(z, s) \\ &= \sum_v \sum_{i,j=1}^{N_v} ((I - z \lambda_v L)^{-1})_{i,j} h_i(\dot{0}) \mu_j(\chi) + C(s, z) \\ &= \sum_v \frac{c^v(z, s)}{(1 - z \lambda_v(s))^{N_v}} + C(s, z) \end{aligned}$$

for some analytic functions  $C(s, z), c^v(s, z)$ .

By taking derivatives we get that

$$\partial_2 \beta(z, 0) = \sum_v \frac{c^v(z, 0)}{(1 - z \lambda_v)^{N_v+1}} + \frac{\partial_2 c^v(z, 0)}{(1 - z \lambda_v)^{N_v}} + \partial_2 C(z, 0).$$

Using the expansion

$$\frac{1}{(1 - z \lambda_v)^{N_v+1}} = \sum_{n=0}^{\infty} \left( \frac{(n + N_v)!}{n! N_v!} \lambda_v^n \right) z^n$$



we see that

$$\begin{aligned}\partial_2\beta(z, 0) &= \sum_{n=1}^{\infty} z^n \sum_{y \in S_n} f^n(y\dot{0}) \\ &= \sum_{n=0}^{\infty} z^n \left( \sum_v D(z) \frac{(n + N_v)!}{n!N_v!} (\lambda_v)^n \right) + E(z)\end{aligned}$$

where  $D(z), E(z)$  are analytic in a neighbourhood of  $z = \frac{1}{\lambda_v}$ , and  $d_v = D(\frac{1}{\lambda_v}) \neq 0$ . By these two expansions, we see that

$$(9.3) \quad M(n) = \sum_{y \in S_n} f^n(y\dot{0}) \sim \sum_v (\lambda_v)^n \left( d_v \frac{(n + N_v)!}{n!N_v!} \right).$$

If we divide both sides of (9.3) by  $n$  then we get the asymptotic

$$(9.4) \quad \sum_{y \in S_n} \frac{f^n(y\dot{0})}{n} \sim \sum_v (\lambda_v)^n \left( \frac{d_v}{n} \frac{(n + N_v)!}{n!N_v!} \right) \sim \sum_v (\lambda_v)^n n^{(N_v-1)} \frac{d_v}{N_v!}$$

as  $n \rightarrow +\infty$ . This corresponds to the statement of the proposition.  $\square$

*Applications to co-compact Fuchsian groups.* We now want to consider the special case of gregarious automata, and in particular the case of co-compact Fuchsian groups. In this particular setting, we are able to deduce a simpler expression for the average  $\mathcal{A}$  of Theorem 1 in the special case where the weighting  $w(g)$  is word length  $|g|$ . We can then consider the dependence of the constant  $\mathcal{A}$  on the underlying geometry.

In the present setting the identity (8.2) takes the simpler form

$$(9.5) \quad (L_{-sr+zf}^n \chi)(\dot{0}) = \lambda(s, z)^{n-1} (\mathbb{P}(s, z) L_{-sr+zf} \chi)(\dot{0}) + (Q(s, z) L_{-sr+zf}^n \chi)(\dot{0})$$

since  $L_{-\delta r}$  has a simple maximal eigenvalue  $\lambda(\delta, 0) = 1$  with eigenprojection  $\mathbb{P}(\delta, 0)$ . Substituting this expression into (8.1) gives us

$$\begin{aligned}(9.6) \quad \eta(s, z) &= \frac{(\mathbb{P}(s, z) L_{-sr+zf} \chi)(\dot{0})}{1 - \lambda(s, z)} + \sum_{n=1}^{\infty} (Q(s, z) L_{-sr+zf}^n \chi)(\dot{0}) \\ &= \frac{1}{1 - \lambda(s, z)} G(s, z) + U_1(s, z)\end{aligned}$$

where  $G(s, z), U(s, z)$  are analytic. Differentiating this last identity gives

$$\begin{aligned}\xi(s) &:= \partial_2 \eta(s, 0) \\ &= \frac{1}{1 - \lambda(s, 0)} \partial_2 G(s, 0) + \frac{\partial_2 \lambda(s, 0)}{(1 - \lambda(s, 0))^2} G(s, 0) + \partial_2 U_1(s, 0)\end{aligned}$$

and observing that

1.  $\partial_2 \lambda(\delta, 0) = 1$  when  $f = 1$ ,
2.  $\partial_1 \lambda(\delta, 0) = -\int r d\mu$ , where  $\mu$  is the equilibrium state of  $-\delta r$ .

we see that

$$\xi(s) = \frac{1}{(\int r d\mu)^2} \frac{1}{(s - \delta)^2} (\mathbb{P} L_{-\delta r} \chi)(\dot{0})$$

Repeating the proof of Theorem 1 with the above form of  $\xi(s)$  we get the asymptotic

$$\rho(T) \sim \frac{1}{(\int r d\mu)^2} \frac{(\mathbb{P}L_{-\delta r}\chi)(\dot{0})}{6} e^{\delta T}$$

Applying similar arguments to  $\eta(s)$  we obtain that

$$\text{Card}\{g \in \Gamma : d(x, gx) \leq T\} \sim \frac{(\mathbb{P}L_{\delta r}\chi)(\dot{0})}{\int r d\mu} e^{\delta T}.$$

In particular, we conclude that the average  $\mathcal{A} = \frac{1}{6 \int r d\mu}$ .

## 10. ASYMPTOTIC AVERAGES BY WORD LENGTH

In section 4 we gave an asymptotic average for the ratio  $\frac{|g|}{d(x, gx)}$ , where the elements  $g \in \Gamma$  are ordered by displacement. In order to state the corresponding result for word length (where word length and displacement are interchanged) we need an additional hypothesis. Specifically, we shall want to assume that the pair  $(\Gamma, \Gamma_0)$  is gregarious.

We recall Cannon's estimate on  $N(n) = \text{Card}\{g \in \Gamma : |g| = n\}$  that there exist constants  $\delta_i$ ,  $D_i$  and positive integers  $l_i$  (for  $i = 1, \dots, M$ ) such that

$$(10.1) \quad N(n) = \sum_{i=1}^M D_i n^{l_i} \delta_i^n \quad \text{as } n \rightarrow +\infty.$$

We can compare this with the result of Proposition 2 that there exists constants  $C_i > 0$ ,  $\beta_i$  and positive integers  $k_i$  ( $i = 1, \dots, m$ ) such that

$$\sum_{\{g \in \Gamma : |g|=n\}} \frac{d(x, gx)}{|g|} \sim \sum_{i=1}^m C_i n^{k_i} \beta_i^n.$$

In the case where we are dealing with a gregarious automaton, we know that the associated subshift of finite type has a single dominant topologically mixing component. In particular, the operator  $L_{-sf}$  has a single simple maximal eigenvalue  $\lambda(s)$  for  $s$  close to 0.

The proof of Proposition 2 becomes correspondingly simplified. Specifically, the asymptotic in (8.3) becomes simplified to

$$\frac{M(n)}{n} = \sum_{z \in S_n} \frac{f^n(z\dot{0})}{n} \sim \left( \frac{d}{ds} (\mathbb{P}(s)(L_{sf}\chi)(\dot{0})) \Big|_{s=0} \right) \lambda^n$$

where we denote  $\lambda(0) = \lambda$  which is the maximal eigenvalue for  $L_0$  (and it is clear that  $\lambda$  is therefore also the unique maximal eigenvalue for the transition matrix  $A$ ).

In Cannon's proof of the rationality of the generating function for word length and thus the corresponding formula for the asymptotic growth of the number  $N(n)$  of group elements of word length  $n$ , the transition matrix  $A$  also determines the asymptotic. Specifically,  $N(n) \sim C\lambda^n$ , for some  $C > 0$ , as  $n \rightarrow +\infty$ .

Comparing these two expressions, we see that the following limit exists

$$(10.2) \quad \lim_{n \rightarrow +\infty} \frac{\sum_{z \in S_n} \frac{f^n(z\dot{0})}{n}}{N(n)} = \frac{\left( \frac{d}{ds} (\mathbb{P}(s)(L_{sf}\chi)(\dot{0})) \Big|_{s=0} \right)}{C}.$$

In particular, we have the following result.

**Proposition 8.** *For gregarious pairs  $(\Gamma, \Gamma_0)$  the averages  $\frac{M(n)}{nN(n)}$  have a limit as  $n \rightarrow +\infty$ .*

## REFERENCES

1. R.Adler and L.Flatto, *Geodesic flows, interval maps and symbolic dynamics*, Bull. Amer. Math. Soc. **25** (1991), 229–334. MR **92b**:58172
2. S. Agmon, *Complex Variable Tauberians*, Trans. Amer. Math. Soc. **74** (1953), 444–481. MR **14**:869a
3. L. Bers, *Uniformization, moduli and Kleinian groups*, Bull. London Math. Soc. **4** (1972), 257–300. MR **52**:5955
4. M. Bourdon, *Actions quasi-convexes d'un groupe hyperbolique, flot géodésique*, PhD thesis (Orsay) 1993.
5. J.Cannon, *The combinatorial structure of co-compact discrete hyperbolic groups*, Geometriae Dedicata **16** (1984), 123–148. MR **86j**:20032
6. M. Coornaert, T. Delzant and A. Papadopoulos, *Géométrie et théorie des groupes*, Lecture Notes in Mathematics, vol. 1441, Springer, Berlin, 1990. MR **92f**:57003
7. H.Delange, *Généralisation du théorème de Ikehara*, Annales Scientifique de Ecole Normale Supérieure **71** (1954), 213–244. MR **16**:921e
8. E. Ghys and P. de la Harpe, *Sur les groupes hyperboliques d'après Mikhael Gromov*, Birkhauser, Boston, 1990. MR **92f**:53050
9. M. Gromov, *Hyperbolic groups*, Essays in group theory, M.S.R.I. publication vol. 8, MSRI, Berkeley, 1987. MR **89e**:20070
10. T. Kato, *Perturbation theory of linear operators*, Springer, 1966. MR **34**:3324
11. S. Lalley, *Renewal theorems in symbolic dynamics, with applications to geodesic flows, non-euclidean tessellations and their fractal limits*, Acta Math. **163** (1989), 1–55. MR **91c**:58112
12. R. Mazzeo and R. Melrose, *Meromorphic extension of the resolvent on complete spaces of asymptotically constant negative curvature*, J. Funct. Anal. **75** (1987), 260–310. MR **89c**:58133
13. J. Milnor, *A note on curvature and the fundamental group*, J. Diff. Geom. **2** (1967), 1–7. MR **38**:636
14. R. Nussbaum, *The radius of the essential spectrum*, Duke Math. J. **37** (1970), 473–478. MR **41**:9028
15. W. Parry and M. Pollicott, *Zeta functions and the periodic orbit structure of hyperbolic dynamics*, Asterisque **187-88** (1990), 1–268. MR **92f**:58141
16. S. Patterson, *On a lattice-point problem in hyperbolic space and related questions in spectral theory*, Arkiv för Mat. **26** (1988), 167–172. MR **89g**:11093
17. M. Pollicott and R. Sharp, *Orbit counting for some discrete groups acting on simply connected manifolds with negative curvature*, Invent. Math. **117** (1994), 275–302. MR **96a**:58147
18. D. Ruelle, *Thermodynamic Formalism*, Addison Wesley, New York, 1978. MR **80g**:82017
19. E. Seneta, *Non-negative matrices*, George Allen and Unwin, London, 1973. MR **52**:10773
20. C. Series, *The infinite word problem and limit sets of fuchsian groups*, Ergod. Th. and Dynam. Sys. **1** (1981), 337–360. MR **84d**:30084
21. C. Series, *Geometrical Markov coding of geodesics on surfaces of constant negative curvature*, Ergod. Th. and Dynam. Sys. **6** (1986), 601–625. MR **88k**:58130
22. D. Sullivan, *The density at infinity of a discrete group of hyperbolic motions*, Publ. Math. (IHES) **50** (1979), 171–202. MR **81b**:58031

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF WARWICK, COVENTRY, CV4 7AL, U.K.

*Current address:* Department of Mathematics, University of Manchester, Oxford Road, Manchester, M13 9PL, U.K.

*E-mail address:* mp@ma.man.ac.uk

MATHEMATICAL INSTITUTE, 24-29 ST. GILES, OXFORD, OX1 3LB, U.K.

*Current address:* Department of Mathematics, University of Manchester, Oxford Road, Manchester, M13 9PL, U.K.

*E-mail address:* sharp@ma.man.ac.uk