

Amenability of linear-activity automaton groups

Gideon Amir*

Omer Angel^{*†}

Bálint Virág^{*‡}

April 2009

Abstract

We prove that every linear-activity automaton group is amenable. The proof is based on showing that a sufficiently symmetric random walk on a specially constructed degree 1 automaton group — the mother group — has asymptotic entropy 0. Our result answers an open question by Nekrashevich in the Kourovka notebook, and gives a partial answer to a question of Sidki.

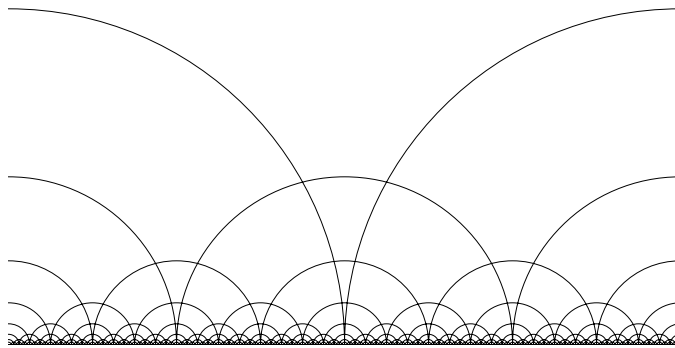


Figure 1: A Schreier graph for a linear-activity group

1 Introduction

Automaton groups are the core of the algebraic theory of fractals. Just as fractals do in geometry, automata groups form a rich new world within group theory. This world has been a source of many interesting examples, but until recently it resisted a general theory. Yet the

*University of Toronto

†University of British Columbia, Supported in part by NSERC of Canada

‡Research supported in part by NSF grant #DMS-0206781.

simplicity of the definitions and the richness of examples suggest the existence of such a theory. The goal of this paper is to make a step in this direction by proving that all linear-activity groups are amenable.

Automaton groups arise in various areas of mathematics; examples include:

The Grigorchuk group, (see Grigorchuk, 1984) which has faster than polynomial but subexponential growth, answering the question of Milnor (1968) about the existence of such groups.

The basilica group, a finitely generated amenable group containing a non-cyclic free semigroup. Its level Schreier graphs approximate the Basilica fractal, that is the Julia set of the polynomial $z^2 - 1$, see Grigorchuk and Żuk (2002) and Bartholdi and Virág (2005).

The Hanoi towers group, the group of possible moves in the Hanoi Towers game on three pegs, a game introduced by Édouard Lucas in 1883. Its level Schreier graphs are discrete Sierpinski gaskets (see Grigorchuk and Šunić, 2006).

The long-range group, an interesting group whose Schreier graphs (Figure) were studied by Benjamini and Hoffman (2005) in the context of long-range percolation theory.

The lamplighter group on \mathbb{Z} , the first example of a group with a Cayley graph that has discrete spectrum, providing also a counterexample to the strong Atiyah conjecture. See Grigorchuk and Żuk (2001).

Automaton groups are subgroups of the automorphism group $\text{Aut}(\mathbb{T}_m)$ of the rooted infinite m -ary tree. Every element of $\text{Aut}(\mathbb{T}_m)$ can be decomposed as

$$g = \langle\langle g_0, \dots, g_{m-1} \rangle\rangle \sigma, \quad g_i \in \text{Aut}(\mathbb{T}_m), \quad \sigma \in \text{Sym}(m),$$

where the g_i , called **first-level sections**, now act on the subtrees rooted at the children of the root of \mathbb{T}_m , and σ permutes these subtrees. An **automaton** A is a finite subset of $\text{Aut}(\mathbb{T}_m)$ so that first level sections of elements of A are also in A . An **automaton group** is a group generated by an automaton. See, e.g., Sunic and Grigorchuk (2007) and references there for a general survey on automaton groups.

A systematic study of automaton groups was initiated by Sidki (2000), who introduced the concept of **activity growth** (or more briefly, activity), a measure of size for automaton groups. We defer the precise definition for later, but note that the activity can be polynomial of any degree or exponential. There are exponential activity growth automaton groups that are isomorphic to the free group (see Glasner and Mozes, 2005; Vorobets and Vorobets, 2007). However, one expects polynomial automaton groups to be smaller. In particular, in contrast with all known finitely generated non-amenable groups, Sidki (2000) showed that polynomial

automaton groups have no free subgroups. In light of this dichotomy, Sidki asked the following natural question.

Question. Are all polynomial activity automaton groups amenable?

It seems that the structure of polynomial automaton groups depends a lot on the degree, and we do not have a conjectured answer to Sidki’s question. The first result in this direction is due to Bartholdi, Kaimanovich, and Nekrashevych (2008), who showed that all degree 0 (also called bounded) automaton groups are amenable. Their proof uses a variant of the self-similar random walk idea introduced in Bartholdi and Virág (2005) to prove the amenability of the basilica group and later streamlined and generalized in Kaimanovich (2005). The goal of this paper is to show the following.

Theorem 1. *All linear-activity automaton groups are amenable.*

Example. A simple example of our theorem is given by the **long-range group**: Let b act on the integers by increasing them by 1, and let a act on integers by increasing them by the lowest power of 2 they are not divisible by. The group generated by a, b is called the long-range group; its Schreier graph is shown in Figure . Its automaton has two states defined recursively by

$$a = \langle\langle a, b \rangle\rangle, \quad b = \langle\langle b, 1 \rangle\rangle(01).$$

The question of the amenability of the long-range group was posed by Nekrashevich, in the Kourovka notebook (2006), Question 16.74, and also in Guido’s book of conjectures (2008), Conjecture 35.9.

Our paper builds on previous work in the theory, including Bartholdi and Virág (2005), Kaimanovich (2005), and Bartholdi et al. (2008). The first part of the proof is to construct a family of linear-activity automaton groups — the **mother groups**— which are then shown to contain subgroups isomorphic to every linear-activity automaton group. This has been done in Bartholdi et al. (2008) for bounded automaton groups. We show how to perform this step for polynomial automata of any degree. It then suffices to show that the linear-activity mother groups are amenable.

The second part is to find a random walk on the mother group with 0 asymptotic entropy. We can only do this for linear-activity mother groups. In fact, we conjecture that there is a phase transition for this question, namely such walks only exist up to degree 2.

Conjecture 2. *The mother group of degree d has a random walk whose step distribution is supported on a finite generating set and whose asymptotic entropy is zero if and only if $d = 0, 1, 2$.*

For $d = 1$ this follows from this paper. For all $d \geq 2$ this conjecture is still open.

Compared to previous work, the random walks we consider are no longer self-similar or even finitely supported. The role of self-similar walks is played by the evolution of random

walk step measures that we can control. For such control, we need a high level of symmetry for the random walk. However, we believe that this is an artifact of the proof.

Conjecture 3. *Every random walk with boundedly supported step distribution on a polynomial automaton group with degree at most 2 has zero asymptotic entropy.*

This conjecture is open even for $d = 0$.

The main tool for the analysis of asymptotic entropy is study of the so-called **ascension operator**. Consider a random walk with step distribution μ on an automaton group, and let v be the first child of the root of \mathbb{T}_m . The section of the walk at v , looked at the times at which the walk fixes v forms another random walk with step distribution μ' . The map $T : \mu \mapsto \mu'$ is called the ascension operator (see further details in Section 4).

Ascension of random walks first considered in Bartholdi and Virág (2005). The ascension operator in this form appeared in Kaimanovich (2005), where the entropy inequality $H_\infty(\mu) \leq H_\infty(T\mu)$ was also proved. Iterating this inequality, one gets $H_\infty(\mu) \leq H_\infty(T^n\mu)$. In this paper, we will analyze $T^n\mu$ in the case when these measures are not finitely supported and not computable exactly. The following proposition allows us to relate $H_\infty(\mu)$ to the asymptotic entropy of limit points of the sequence of measures $T^n\mu$.

Proposition 4 (Upper semi-continuity of asymptotic entropy). *If ν_n, ν are probability measures on a countable group so that $\nu_n \rightarrow \nu$ weakly and $H(\nu_n) \rightarrow H(\nu)$, then*

$$\limsup_{n \rightarrow \infty} (H_\infty(\nu_n)) \leq H_\infty(\nu).$$

In light of Proposition 4, it suffices to find a subsequence along which $T^n\mu \rightarrow \nu$ and $H(T^n\mu) \rightarrow H(\nu)$, and ν has zero asymptotic entropy. Perhaps surprisingly, it is not too difficult to show that for appropriate μ , any subsequential limit point ν has zero asymptotic entropy. For the other two claims, it suffices to show the tightness and entropy-tightness (defined below) of the sequence $T^n\mu$. Proving these facts takes up a large part of this paper. Showing the tightness of the sequence $T^n\mu$ is the main obstacle to extending our proof to the higher degree case. For convenient reference, we summarize the preceding discussion in a theorem.

Theorem 5 (Asymptotic entropy of automaton groups). *If the group generated by the support of μ acts transitively on all levels of the tree and the sequence $T^k\mu$ is entropy-tight, then for every subsequential weak limit point ν we have*

$$H_\infty(\mu) \leq H_\infty(\nu).$$

The main challenge is to construct measures for which the ascension operator is tractable. The measures that we consider are based on the uniform measures \bar{q}_i on certain finite subgroups of the mother group \mathfrak{M} . They have the property that if a probability measure μ on \mathfrak{M} is a

convex combination of convolution products of the \bar{q}_i 's, then so is $T\mu$. Thus the iterated ascensions $T^n\mu$ of such measures can be understood in terms of the \bar{q}_i 's, which are sufficiently symmetric and can be controlled. Further details can be found in Section 6.

Organization. The structure of this paper is as follows. Sections 2–4 contain the setup, review and proof of some preliminary results; readers familiar with previous work on the subject may want to skip these sections. Section 2 reviews basic concepts related to automata groups. In Section 3 we do the same for entropy, and prove Proposition 4. In Section 4 we introduce and study the ascension operator. Mother groups are defined in Section 5, and it is shown that they contain all polynomial automata groups. In Section 6 we introduce a special class of measures, called patterns, and an algebraic way to study them. We also define an ascension operator for patterns. In Section 7 we study properties of iterated ascension on patterns. Section 8 contains some preliminary results on entropy of pattern measures, and finally, the main theorem is proved in 9.

2 Automata and their groups

Basic definitions. Finite automata are the simplest interesting model of computing; we first connect our definition of automata to the more traditional one.

The space of words in alphabet $\{0, \dots, m-1\}$ has a natural tree structure, with $\{wx\}_{x < m}$ being the children of the finite word w , and the empty word \emptyset being the root. Let \mathbb{T}_m denote this tree.

A **finite automaton** on m symbols is a finite set of states A with a map $A \rightarrow A^m \times \text{Sym}(m)$ sending $a \mapsto (a_0, \dots, a_{m-1}, \sigma_a)$. We will use the notation

$$a = \langle\langle a_0, \dots, a_{m-1} \rangle\rangle \sigma_a.$$

An automaton acts on words in alphabet $\{0, \dots, m-1\}$ sequentially. When the automaton is in a state a and reads a letter x , it outputs $x.\sigma_a$ and moves to state a_x . From this state the automaton acts on the rest of the word. Symbolically, for a word xw (starting with a letter x) we have the recursive definition

$$(xw).a = (x.\sigma_a)(w.a_x). \tag{1}$$

The first k symbols of the output are determined by the first k symbols read. Note that the action is defined for both finite and infinite words, and that the action on infinite words determines the action on finite words and vice versa. It follows that each element $a \in A$ is an automorphism of \mathbb{T}_m . The **automaton group** corresponding to an automaton A is the subgroup of $\text{Aut}(\mathbb{T}_m)$ generated by A .

The action (1) corresponds to the following **multiplication rule**:

$$\langle\langle a_0, \dots, a_d \rangle\rangle \sigma \langle\langle b_0, \dots, b_d \rangle\rangle \tau = \langle\langle a_0 b_{0,\sigma}, \dots, a_{m-1} b_{(m-1),\sigma} \rangle\rangle \sigma \tau.$$

This multiplication rule can be used to define automaton groups without any reference to automorphisms of the tree. However, keeping the action on the tree in mind makes some constructions used in the proof more natural.

We use the conjugation notation $a^b = b^{-1}ab$.

The notion of first-level sections can be generalized to any level. If $v \in \mathbb{T}_m$ is a finite word and $g \in \text{Aut}(\mathbb{T}_m)$, then there is a word v' of equal length to v and an automorphism $g' \in \text{Aut}(\mathbb{T}_m)$ such that $vw.g = v'(w.g')$, for every word w . This g' is called the **section** of g at v . Informally, g' is the action of g on the subtree above the vertex v . The section of g at v is denoted $g(v)$.

Activity growth of automaton groups. For any state $a \in A$, the number of length- n words v such that the section $a(v)$ is not the identity satisfies a certain linear recursion. Thus this number grows either polynomially with some degree d or exponentially. We define the **degree of activity growth** (in short, degree) of a to be d or ∞ , respectively. The **degree** of an automaton group is the maximal degree of any of its generators. Automaton groups are said to have **bounded, linear, polynomial** or **exponential** activity growth when their degree is 0, 1, finite or infinite, respectively.

Degree and cycle structure. An automaton gives rise to a directed graph where there is a directed edge from a state a to each of its first-level sections $a(i)$. If the same state appears more than once as an a_i then there are parallel edges. If a appears as a first-level section of itself (i.e., $a(i) = a$) then there is a loop at a . Thus the only edges leaving the identity (if it is a state of A) are loops back to the identity. The loops at the identity are called **trivial cycles**. To avoid degeneracies in the graph, we will assume from now on that all non-identity states in the automaton act non-trivially. This assumption, which does not limit the generality of the automaton groups considered, allows some connections between the structure of the said graph and properties of the automaton.

The number of active vertices of a at level n is just the number of directed paths of length n starting at a and ending anywhere but the identity. It follows easily that an automaton is exponential if and only if there are two nontrivial cycles so that each is reachable from the other via a directed path.

The nontrivial directed cycles in the directed graph of a polynomial automaton have a partial order: $c_1 < c_2$ if $c_1 \neq c_2$ and there is a directed path from a state of c_2 to a state of c_1 . Define the degree of a cycle c as the maximal n so that there is an increasing chain $c_0 < c_1 < \dots < c_n = c$. It is straightforward to see that the degree of a polynomial automaton

is the maximal degree of any cycle in its directed graph. The degree of a state g is the maximal degree of a cycle reachable via a directed path from g . The identity 1 is always considered to have degree -1 .

Hierarchy levels. For polynomial automaton, the hierarchy level of a state is a refinement of its degree, taking values in the sequence

$$-\tilde{1}, -1, \tilde{0}, 0, \tilde{1}, 1, \tilde{2}, 2, \dots$$

States in cycles of degree d have **hierarchy level** \tilde{d} . States of degree d that are not in a cycle have hierarchy level d . A state can point to states either in its own hierarchy level or lower levels. The only state with hierarchy level $-\tilde{1}$ is the identity.

Collapsing levels. By dividing words in m symbols into blocks of length k , we can view them as words in m^k symbols. Similarly, given an automaton A acting on m symbols, it naturally gives rise to an automaton with the same states acting on words in m^k symbols; we call this new automaton the k -**collapse** of A , and it acts on the tree \mathbb{T}_{m^k} .

Claim 2.1. *The automaton groups of A and of its k -collapse are isomorphic.*

Proof. The key is that the vertices of \mathbb{T}_{m^k} are naturally associated with every k 'th level of \mathbb{T}_m , and that $\text{Aut}(\mathbb{T}_m) \subset \text{Aut}(\mathbb{T}_{m^k})$. Restricting this embedding to $G = \langle A \rangle$ gives an isomorphism from the automaton group of A to the group of the k -collapse of A . \square

3 Entropy and asymptotic entropy

The purpose of this section is to review the notion of entropy and to prove Proposition 3.3 below, which gives a condition for upper semi-continuity of the asymptotic entropy of a random walk on a group.

Recall that the **entropy** of a finite non-negative measure μ supported on a countable set G is defined by

$$H(\mu) = \sum_{x \in G} -\mu(x) \log \mu(x),$$

where by convention $0 \log 0 = 0$. The entropy $H(X)$ of a discrete random variable X is given by the entropy of its distribution; the entropy $H(X_1, \dots, X_n)$ of more random variables is given by the entropy of the joint distribution of the X_i . In order to define conditional entropy for two random variables X, Y , let $f(y)$ denote entropy of the conditional distribution of X given $Y = y$. Then the **conditional entropy of X given Y** is defined as $H(X|Y) := \mathbb{E}f(Y)$.

The conditional entropy satisfies

$$H(X, Y) = H(X|Y) + H(Y).$$

A useful and easy fact is that among measures supported on a given finite set, the one having maximal entropy is the uniform measure on that set. Thus for a measure μ supported on a finite set B we have $H(\mu) \leq \log(|B|)$. It is not hard to construct measures on \mathbb{Z} with finite expectation and infinite entropy, but it is known (see, for example, Kemp, 1997) that among integer-valued random variables with given mean and variance the discrete normal distribution maximizes entropy. In particular, the following holds.

Lemma 3.1. *There is a constant c such that for any integer-valued random variable N we have*

$$H(N) \leq C + (\log \text{Var } N)^+,$$

where $x^+ = \max(0, x)$.

Define the **asymptotic entropy** of a sequence of random variables X_n as

$$\limsup_{n \rightarrow \infty} \frac{1}{n} H(X_n).$$

If a random walk on a group G has i.i.d. steps with distribution given by μ , its asymptotic entropy is given by

$$H_\infty(\mu) = \lim_{n \rightarrow \infty} \frac{1}{n} H(\mu^{*n}),$$

where μ^{*n} is the n -fold convolution of μ , or, equivalently, the distribution of the n 'th step of the random walk. The limit exists by subadditivity.

We say that a set of probability measures $\{\mu_i\}$ is **entropy-tight** if for every ε there exists a finite set K so that for all i

$$\sum_{x \notin K} -\mu_i(x) \log \mu_i(x) < \varepsilon. \tag{2}$$

In other words, entropy-tightness means the uniform integrability of the function $\mu_i(x) \log \mu_i(x)$ with respect to counting measure. The importance of entropy-tightness comes from the following direct application of Vitali's convergence lemma.

Lemma 3.2. *Assume that $\mu_n \rightarrow \mu$ weakly. Then $H(\mu_n) \rightarrow H(\mu)$ if and only if the sequence μ_n is entropy-tight.*

The aim of this section is the following result:

Proposition 3.3. *If $\mu_i \rightarrow \mu$ and $\{\mu_i\}$ is entropy-tight, then $\limsup H_\infty(\mu_i) \leq H_\infty(\mu)$.*

We begin with two lemmas.

Lemma 3.4. For any two measures μ and ν we have $H(\mu * \nu) \leq |\nu|H(\mu) + |\mu|H(\nu)$.

Proof. For the product measure $\mu \times \nu$ we have

$$\begin{aligned} H(\mu \times \nu) &= \sum_{x,y} \mu(x)\nu(y) \log(\mu(x)\nu(y)) \\ &= \sum_{x,y} \mu(x)\nu(y) \log \mu(x) + \sum_{x,y} \mu(x)\nu(y) \log \nu(y) \\ &= |\mu|H(\nu) + |\nu|H(\mu). \end{aligned}$$

Now by subadditivity of the function $-x \log x$ we have $H(\mu * \nu) \leq H(\mu \times \nu)$. The lemma follows. \square

Lemma 3.5. If a family of probability measures $\{\mu_i\}$ is both tight and entropy-tight then so is the set of all their n -fold convolutions $\{\mu_{i_1} * \dots * \mu_{i_n}\}$.

Proof. It suffices to prove this for $n = 2$, with larger n following by induction. For a measure μ we denote the measure restricted to A , without normalization, by μ^A .

Fix some $\varepsilon > 0$. The conditions imply that there is a finite set K such that for all i ,

$$|\mu_i^{K^c}| = \mu_i(K^c) < \varepsilon, \quad H(\mu_i^{K^c}) < \varepsilon.$$

For any two measures μ, ν from the set we have

$$\mu * \nu = \mu^K * \nu^K + \mu^{K^c} * \nu^K + \mu^K * \nu^{K^c} + \mu^{K^c} * \nu^{K^c}.$$

Consider the set finite set $B = K * K$, and note that the support of $\mu^K * \nu^K$ is contained in B . It follows that

$$|(\mu * \nu)^{B^c}| \leq |\mu^{K^c} * \nu^K + \mu^K * \nu^{K^c} + \mu^{K^c} * \nu^{K^c}| < 2\varepsilon + \varepsilon^2,$$

which can be made arbitrarily small, hence the convolutions form a tight family.

Note that for $a, b \geq 0$ we have $-a \log a - b \log b > -(a + b) \log(a + b)$, so that entropy of measures is subadditive. Note also that entropy-tightness implies that for some M , for every i we have $H(\mu_i) \leq M$.

It follows using Lemma 3.4 that

$$\begin{aligned} H((\mu * \nu)^{B^c}) &\leq H(\mu^{K^c} * \nu^K + \mu^K * \nu^{K^c} + \mu^{K^c} * \nu^{K^c}) \\ &< 2M\varepsilon + 2\varepsilon + 2\varepsilon^2. \end{aligned}$$

Since this too can be made arbitrarily small, the convolutions are also entropy-tight. \square

Proof of Proposition 3.3. Fix some n . By Lemma 3.5 we have that $\{\mu_i^{*n}\}_i$ are entropy-tight, and so by Lemma 3.2 we have

$$H(\mu^{*n}) = \lim_{i \rightarrow \infty} H(\mu_i^{*n}).$$

The sequence $H(\mu_i^{*n})$ is subadditive, and therefore

$$H_\infty(\mu_i) \leq \frac{1}{n} H(\mu_i^{*n}).$$

Taking the lim sup of both sides we get

$$\limsup_{i \rightarrow \infty} H_\infty(\mu_i) \leq \frac{1}{n} \limsup_{i \rightarrow \infty} H(\mu_i^{*n}) = \frac{1}{n} H(\mu^{*n}).$$

Since n is arbitrary, we can take a limit as $n \rightarrow \infty$ to conclude the proof. \square

4 Random walks and the ascension operator

In our proof we use the method, introduced by Bartholdi and Virág (2005), of studying a random walk on an automaton group by looking at its induced action on the subtree above a vertex v , specifically at times n at which X_n fixes v .

Random walks on quotients and subgroups. Let $\{X_n\}_{n \geq 0}$ be a random walk on a countable group G started at $X_0 = 1$.

If N is a normal subgroup of G , and $K = G/N$, then X has a natural projection to K , namely $Y_n = NX_n$. We call this the **quotient random walk** on G/N .

A set $S \subset G$ is called recurrent (for a random walk X) if X visits S infinitely often with probability one. For example, every finite index subgroup is recurrent in G . For a recurrent subgroup S the steps at which the random walk visits S form a random walk Y on S (i.e. if τ_n is the n th visit to S , let $Y_n = X_{\tau_n}$). We call this the **induced random walk on S** .

The ascension operator. Given an automaton group A and a vertex $j \in \mathbb{T}_m$, consider the stabilizer subgroup A_j of j . Consider also the subgroup A'_j which is the stabilizer of the entire subtree above vertex j . Then A'_j is normal in A_j . For $g \in A_j$ the coset gA'_j consists of all elements with the same action $g(j)$ on the subtree above j . Since A is an automaton group, $g(j) \in A$, hence the group A_j/A'_j is therefore canonically isomorphic to a subgroup of A .

We now specialize to vertex $j = 0$ in the first level of \mathbb{T}_m . Given a random walk on A with step distribution μ , we can consider the induced walk on A_0 , and then its quotient walk on A_0/A'_0 . By the above, this again can be viewed as a random walk on A with step distribution μ' . The **ascension operator** T is the operator that maps each probability measure μ on A , to the measure μ' above. If τ_n are the times at which X_{τ_n} fixes vertex 0, then the actions of X_{τ_n} on the subtree above 0 are a random walk with step distribution $T\mu$.

We say that a measure μ is transitive on level k if the group generated by its support acts transitively on that level of \mathbb{T}_m . We will use the following entropy inequality.

Theorem 4.1 (Kaimanovich (2005), Theorem 3.1). *Assume that a probability measure μ is transitive on the first level. Then the asymptotic entropies satisfy $H_\infty(\mu) \leq H_\infty(T\mu)$.*

The upper semi-continuity of asymptotic entropy, Proposition 4 yields the following.

Theorem 4.2 (Asymptotic entropy of automaton groups). *If the group generated by the support of μ is transitive on all levels and the sequence $\{T^k\mu\}$ is entropy-tight, then for any subsequential limit point ν we have*

$$H_\infty(\mu) \leq H_\infty(\nu).$$

Proof. The transitivity of μ implies that $T^k\mu$ is transitive on the first level for all k . Repeated application of Theorem 4.1 shows that for each k

$$H_\infty(\mu) \leq H_\infty(T^k\mu).$$

Taking limsup along the subsequence converging to ν and using Proposition 3.3 gives

$$H_\infty(\mu) \leq \limsup H_\infty(T^k\mu) \leq H_\infty(\nu). \quad \square$$

5 Mother groups

The (m, r) -**mother group**, denoted $\mathfrak{M}_{m,r}$, is defined as the automaton group generated by the following states

$$\begin{aligned} \alpha_{k,\sigma} &= \langle\langle \alpha_{k,\sigma}, \alpha_{k-1,\sigma}, 1, \dots, 1 \rangle\rangle, & 0 \leq k \leq r, \\ \alpha_{-1,\sigma} &= \sigma \\ \beta_{k,\rho} &= \langle\langle \beta_{k,\rho}, \beta_{k-1,\rho}, 1, \dots, 1 \rangle\rangle, & 1 \leq k \leq r, \\ \beta_{0,\rho} &= \langle\langle \beta_{0,\rho}, 1, \dots, 1 \rangle\rangle\rho. \end{aligned}$$

where $\sigma, \rho \in \text{Sym}(m)$ are arbitrary, subject to $0.\rho = 0$. The number of states in $\mathfrak{M}_{m,r}$ as defined here is $m!(r+1) + (m-1)!r$. The same group can be generated with fewer states by taking σ, ρ only in a minimal, (2-element) set of generators of $\text{Sym}(m)$ and $\text{stab}(0) \subset \text{Sym}(m)$ respectively. This would give a generating set of size $4r+2$. After this section m will not play an important role, so we also denote the group by \mathfrak{M}_r .

The actions of $\alpha_{k,\sigma}$ and $\beta_{k,\rho}$ on a word have simple descriptions. Both read the word and make no changes up to the $k+1$ 'th non-zero letter.

- If the first $k+1$ nonzero letters in a word are all 1, then $\alpha_{k,\sigma}$ permutes the next letter by σ . Otherwise it does nothing.

- If the first k nonzero letters in a word are all 1, then $\beta_{k,\rho}$ permutes the next nonzero letter by ρ . Otherwise it does nothing.

Thus both affect only the $k + 1$ 'st non-zero letter and the letter immediately following it.

Note that $\alpha_{k,\sigma}$ and $\beta_{k,\rho}$ both have self-loops and are of degree k , so they have hierarchy level \tilde{k} .

Theorem 5.1 (Mother groups contain all). *Every degree- r automaton group is isomorphic to a subgroup of $\mathfrak{M}_{m,r}$ for some m .*

Note that m is generally not the same as the degree of the tree on which the automaton acts.

Proof. We prove that there exist m', m'' so that each of the following groups can be isomorphically embedded in the next: $A \subset G_{m',r} \subset \mathfrak{M}_{m',r}^* \subset \mathfrak{M}_{m'',r}$. The intermediate groups are defined below. The three containments are proved in Lemmas 5.2, 5.6, 5.7. \square

The first reduction

For each hierarchy level $h \in \{-\tilde{1}, -1, \tilde{0}, 0, \tilde{1}, 1, \tilde{2}, 2, \dots\}$ we define an automaton group as follows. $G_{m,-\tilde{1}} = \{1\}$, and $G_{m,-1} = \text{Sym}(m)$. For any $r \geq 0$ the group $G_{m,\tilde{r}}$ is obtained by adding to $G_{m,r-1}$ all possible states of the form:

$$g = \langle\langle g_0, \dots, g_{m-1} \rangle\rangle \sigma$$

(with $\sigma \in \text{Sym}(m)$) satisfying the following. There is a unique i such that $g_i = g$, and for $j \neq i$ we have that g_j is a state of $G_{m,r-1}$.

Similarly, for $r \geq 0$ we define $G_{m,r}$ by adding to $G_{m,\tilde{r}}$ all possible states of the form:

$$g = \langle\langle g_0, \dots, g_{m-1} \rangle\rangle \sigma$$

where $\sigma \in \text{Sym}(m)$, and g_j is a state of $G_{m,\tilde{r}}$ for all j .

Since they are larger than the mother groups, and are also predecessors of the mother groups, in this section we will refer to the groups $G_{m,h}$ as **grand-mother groups**.

Lemma 5.2. *Let A be an automaton group with all states of hierarchy level at most h . Then for some k , the k -collapse of A is isomorphic to a subgroup of $G_{m^k,h}$*

Proof. Let $k \in \mathbb{Z}$ be a multiple of all cycle lengths of A and larger than the length of any simple directed path in A . The k -collapsed version A' of A has the property that all of its cycles are loops. States with loops have a single loop only, since otherwise A would have been exponential.

A state of a given hierarchy level can only point to a state from a lower level, except for states with loops that also point to themselves. By construction, the grandmother group $G_{m',r}$ contains all possible states of this form. \square

Extended mother groups

Let W_k denote the finite subgroup of $\text{Aut}(\mathbb{T}_m)$ for which all sections at level k are the identity.

Definition 5.3. The **extended mother groups** $\mathfrak{M}_{m,\tilde{r}}^*$, $\mathfrak{M}_{m,r}^*$ are defined the same way as the ordinary mother groups $\mathfrak{M}_{m,r}$, except that σ ranges over all elements of W_{r+2} and W_{r+3} , respectively (the options for ρ are unchanged).

The extended mother groups are nested:

$$\mathfrak{M}_{m,k-1}^* \subset \mathfrak{M}_{m,\tilde{k}}^* \subset \mathfrak{M}_{m,k}^*.$$

Moreover, we have the following lemma.

Lemma 5.4. *If $g_0, \dots, g_{m-1} \in \mathfrak{M}_{m,\tilde{r}}^*$, then*

$$g = \langle\langle g_0, \dots, g_{m-1} \rangle\rangle \in \mathfrak{M}_{m,r}^*.$$

Proof. By taking products, it suffices to prove this for the case when all g_i are the identity except for one, which is a generator of $\mathfrak{M}_{m,\tilde{r}}^*$. Moreover, since $\text{Sym}(m) \subset \mathfrak{M}_{m,r}^*$, by conjugating by a transposition $\tau \in \text{Sym}(m)$, we may assume that the non-identity entry is g_0 .

We now prove the claim by induction on the degree of g_0 . Consider first elements of type α . For $g_0 = \alpha_{-1,\sigma} = \sigma$ this holds by the definition of $\mathfrak{M}_{m,r}^*$ (this is the reason for the choice of W_{r+3} in the definition). For higher degree elements, by definition

$$\alpha_{k,\sigma} = \langle\langle \alpha_{k,\sigma}, \alpha_{k-1,\sigma}, 1, \dots, 1 \rangle\rangle,$$

so

$$\langle\langle \alpha_{k,\sigma}, 1, \dots, 1 \rangle\rangle = \alpha_{k,\sigma} \langle\langle 1, \alpha_{k-1,\sigma}, 1, \dots, 1 \rangle\rangle^{-1}.$$

Now, $\alpha_{k,\sigma} \in \mathfrak{M}_{m,\tilde{r}}^* \subset \mathfrak{M}_{m,r}^*$, and the induction hypothesis implies $\langle\langle 1, \alpha_{k-1,\sigma}, 1, \dots, 1 \rangle\rangle \in \mathfrak{M}_{m,r}^*$ as well, hence so is their product.

The proof for type β elements goes the same way. □

Lemma 5.5. *If $g_1, \dots, g_m \in \mathfrak{M}_{m,r-1}^*$, and $\rho \in \text{Sym}(m)$ has $0.\rho = 0$ then*

$$a = \langle\langle a, g_1, \dots, g_m \rangle\rangle \rho \in \mathfrak{M}_{m,\tilde{r}}^*.$$

Proof. If g is a state of $\mathfrak{M}_{m,r-1}^*$ then by definition $a = \langle a, g, 1, \dots, 1 \rangle \in \mathfrak{M}_{m,\tilde{r}}^*$. Multiplying such states shows that this is true for any $g \in \mathfrak{M}_{m,r-1}^*$. Define $a_i = \langle a_i, g_i, 1, \dots \rangle$. We have

$$a = \left(\prod_i a_i^{(1i)} \right) \rho$$

where $(1j)$ is the transposition. (ij) and ρ are in $\mathfrak{M}_{m,\tilde{r}}^*$ since they are of type β_0 . Thus $a \in \mathfrak{M}_{m,\tilde{r}}^*$ as claimed. □

The key step in the following construction is a conjugation due to Bartholdi et al. (2008), where it was used for bounded automaton groups. Consider the automorphism

$$\delta = \langle\langle \delta, \delta\gamma^{-1}, \delta\gamma^{-2}, \dots \rangle\rangle\sigma$$

where γ is the cyclic shift $(012\dots m-1)$.

Lemma 5.6. *For every hierarchy level h the conjugated grandmother group $G_{m,h}^\delta$ is a subgroup of the mother group $\mathfrak{M}_{m,h}^*$.*

Proof. The proof proceeds by induction on the h . On each level, it suffices to show this for the group generators. For hierarchy level -1 we have

$$\sigma^\delta = \langle\langle \gamma^{0-0.\sigma}, \dots, \gamma^{m-1-(m-1).\sigma} \rangle\rangle\sigma$$

is an element of $W_2 = \mathfrak{M}_{m,-1}^*$.

Assume the lemma holds for hierarchy level $r-1$. Let $g = \langle\langle g_0, \dots, g_{m-1} \rangle\rangle\sigma$ be a generator of $G_{m,\tilde{r}}$. Assume that $g_k = g$ is the loop at state g . We need to prove that g^δ is an element of $\mathfrak{M}_{m,\tilde{r}}^*$, for which it suffices to show that this holds for $h = \gamma^k g^\delta \gamma^{-k.\sigma}$. We have

$$h = \langle\langle h_0, \dots, h_{m-1} \rangle\rangle\gamma^k \sigma \gamma^{-k.\sigma}, \quad \text{where} \quad h_i = \gamma^{i+k} g_{i+k}^\delta \gamma^{-(i+k).\sigma}.$$

Now $h_0 = h$, the h_1, h_2, \dots are in $\mathfrak{M}_{m,r-1}^*$, and $\gamma^k \sigma \gamma^{-k.\sigma}$ fixes 0. Thus by Lemma 5.5 we have that $h \in \mathfrak{M}_{m,\tilde{r}}^*$.

To get from hierarchy level \tilde{r} to hierarchy level r , suppose that $g = \langle\langle g_0, \dots, g_{m-1} \rangle\rangle\sigma$ where g_i are generators of $\mathfrak{M}_{m,\tilde{r}}^*$. Then

$$g^\delta = \langle\langle g_0^{\delta\gamma^0}, \dots, g_{m-1}^{\delta\gamma^{m-1}} \rangle\rangle\sigma^\delta \in \mathfrak{M}_{m,r}^*$$

by the inductive hypothesis and Lemma 5.4. □

Embedding in \mathfrak{M}

The extended mother groups contain the ordinary mother groups, but there are also embeddings in the other direction, as the following lemma shows.

Lemma 5.7. *The $r+3$ -collapsed version of $\mathfrak{M}_{m,r}^*$ is isomorphic to a subgroup of $\mathfrak{M}_{m^{r+3},r}$.*

Proof. It suffices to show this for generators. Clearly, the $r+3$ -collapse of any element in W_{r+3} is just an element in $\text{Sym}(m^{r+3})$. The $r+3$ -collapsed version of any other state of type α or β is of the form

$$a = \langle\langle a, g_1, \dots, g_{m'-1} \rangle\rangle\rho$$

where $\rho \in \text{Sym}(m^{r+3})$ and $0.\rho = 0$, and the g_i 's are states of lower degree than a . Just as in Lemma 5.5, we define a_i from g_i by

$$a_i = \langle\langle a_i, g_i, 1, \dots, 1 \rangle\rangle$$

and we note that a_i is a state of $\mathfrak{M}_{m^{r+3}, r}$. Any ρ with $0.\rho = 0$ is just β_0, ρ , and so is also in \mathfrak{M} . Thus we can conjugate by the transposition $(1i)$ and write

$$a = \left(\prod_{i=1}^{m-1} a_i^{(1i)} \right) \rho$$

as an element of $\mathfrak{M}_{m^{r+3}, r}$. □

Level subgroups

Observe that the group of automorphisms of the first two levels of \mathbb{T}_m fixing 0 and its children is isomorphic to $\text{Sym}(m) \wr \text{Sym}(m-1)$.

For each $\sigma \in \text{Sym}(m) \wr \text{Sym}(m-1)$ and each word w in the symbols $\{1, \dots, m-1\}$ let $\lambda_{w, \sigma}$ denote the element of $\text{Aut}(T)$ acting as follows: If the first $|w|$ nonzero letters agree with w , then $\lambda_{w, \sigma}$ permutes the $|w| + 1$ st nonzero letter and the following letter by σ . Otherwise $\lambda_{w, \sigma}$ does nothing.

For a word w of length k , define the group $\mathfrak{L}_{m, k}^w$ generated by $\lambda_{w, \sigma}$ as σ ranges over $\text{Sym}(m) \wr \text{Sym}(m-1)$. Define the group $\mathfrak{L}_{m, k}$ to be the group generated by the $\mathfrak{L}_{m, k}^w$ for all words w of length k . Define further $\mathfrak{L}_{m, -1} = \text{Sym}(m)$.

Later, we will consider random walks on the mother group whose step distribution is a convex combination of uniform measures on the subgroups $\mathfrak{L}_{m, k}$ for various k 's.

Lemma 5.8. *For each w , $\mathfrak{L}_{m, k}^w \approx \text{Sym}(m) \wr \text{Sym}(m-1)$. The group $\mathfrak{L}_{m, k}$ is a subset of $\mathfrak{M}_{m, k}$ and is the direct product of $\mathfrak{L}_{m, k}^w$ for $w \in \{1, \dots, m-1\}^k$.*

Proof. The structure of $\mathfrak{L}_{m, k}^w$ follows from $\lambda_{w, \sigma} \lambda_{w, \sigma'} = \lambda_{w, \sigma \sigma'}$. Generators corresponding to different words w of equal length commute, hence $\mathfrak{L}_{m, k}$ is the direct product of the $\mathfrak{L}_{m, k}^w$'s.

For $w = 11 \dots 1$ of length k we have that $\mathfrak{L}_{m, k}^w$ is generated by the $\beta_{k, \rho}$ and $\alpha_{k, \sigma}$'s, hence it is a subgroup of $\mathfrak{M}_{m, k}$.

Now let w be a general word of length k , and note that the automorphism

$$b = \beta_{k-1, (1w_k)} \cdots \beta_{0, (1w_1)}$$

changes only the first k non-zero letters, and if they are all 1 it changes them to the letters of w . Thus $\mathfrak{L}_{m, k}^w$ is the conjugate by b of $\mathfrak{L}_{m, k}^{11 \dots 1}$, and therefore $\mathfrak{L}_{m, k}^w \subset \mathfrak{M}_{m, k}$ for any w . □

6 Patterns

The random walks on the mother groups that we consider below have a step distribution that is a mixture of the uniform measures \bar{q}_i on the finite level subgroups \mathfrak{L}_i , and their convolutions. It is convenient to think of the measures \bar{q}_i as elements of the group algebra $\mathbb{R}\mathfrak{M}_k$.

The elements $\bar{q}_i \in \mathbb{R}\mathfrak{M}_k$ satisfy the relations $\bar{q}_i^2 = \bar{q}_i$ (being uniform measures on subgroups), and possibly some other relations that are less tractable. Therefore we introduce the more basic semi-group (with identity \emptyset) \mathcal{Q}_k given by the generators q_{-1}, \dots, q_k , (which we call **pattern letters**) and the relations $q_i^2 = q_i$. Elements of \mathcal{Q}_k are called **patterns**. We further define $D = D_k \subset \mathcal{Q}_k$ to be the set of patterns that contain the letter q_{-1} .

Since the measures \bar{q}_i satisfy all relations satisfied by the q_i , the map $q_i \mapsto \bar{q}_i$ extends multiplicatively to a unique semi-group homomorphism $\mathcal{Q}_k \rightarrow \mathbb{R}\mathfrak{M}_k$. The image \bar{p} of a pattern p is called the **evaluation** of p .

Each pattern is an equivalence class of words in the letters $\{q_{-1}, q_0, \dots, q_k\}$. The equivalence relation is that repetition of a symbol is equivalent to a single instance. For example, for a, b, c in the set of letters, we have $abaacaabba \equiv abacaba$. Composition is concatenation. The **length** of a pattern p , i.e. the length of the shortest element of its equivalence class is denoted $|p|$. The set \mathcal{Q}_0 is of particular interest to us. It contains patterns in two letters, which must alternate. Thus \mathcal{Q}_0 has only two patterns of any length. Finally, for a measure μ on patterns and $\alpha \geq 0$ the α -**moment** of the length is denoted by

$$\mu(|p|^\alpha) = \sum_{p \in \mathcal{Q}_k} |p|^\alpha \mu(p).$$

The main reason for the definition of patterns is that they have behave nicely with respect to the ascension operator. To make this statement more precise, consider a probability measure μ supported on patterns. μ is a convex combination of measures concentrated on a single pattern, and is naturally an element of $\mathbb{R}\mathcal{Q}_k$. Its image $\bar{\mu}$ under the quotient map is an element of $\mathbb{R}\mathfrak{M}_k$, or more precisely, a probability measure on \mathfrak{M}_k . The measure $\bar{\mu}$ is will also be called the **evaluation** of μ . It is given by the formula

$$\bar{\mu} = \sum_{p \in \mathcal{Q}_k} \mu(p) \bar{p}.$$

Probability measures of the form $\bar{\mu}$ are special. Nevertheless, this class is preserved by the ascension operator T . This suggests that the ascension operator can be defined on the level of patterns.

The ascension formula

Let $\{Z_n\}$ be the random walk on G , with steps X_n , so that $Z_n = X_1 \cdots X_n$, and write $X_n = \langle\langle X_n(0), \dots, X_n(m-1) \rangle\rangle \sigma_n$. Clearly the trajectory of 0, is itself a Markov chain: If $J_n = 0.Z_n$

then $J_n = J_{n-1} \cdot \sigma_n$. The section $Z_n(0)$ is not itself a Markov chain (except for very restricted step distributions). However, the pair $(J_n, Z_n(0))$ is a Markov chain, since $Z_n(0) = Z_{n-1}(0)X_n(J_{n-1})$: to determine the action of Z_n on the subtree above 0 we need to know the action of Z_{n-1} on that subtree, and a single section of X_n , with index given by $0 \cdot Z_{n-1}$.

A key observation used by Kaimanovich (2005) is that if we consider the process $\{Z_n(0)\}$ conditioned on $\{J_n\}$ we get a process with independent increments. Moreover, if $Z_n(0) = Y_1 \cdots Y_n$, then Y_n depends on $\{J_n\}$ only through the pair J_{n-1}, J_n . Indeed the law of Y_n is the distribution of $Z_n(J_{n-1})$ conditioned on $J_{n-1} \cdot \sigma_n = J_n$.

Such a process can be described naturally by the $m \times m$ matrix M with entries in the group algebra $\mathbb{R}G$, with entries

$$M_{ij} = \mathbb{E}X(i)1_{\{i \cdot X=j\}}. \quad (3)$$

Here $X \in \mathfrak{M}_k$ is a step of the original random walk, and therefore the section $X(i)$ is also in the group. To take its expectation consider $X(i)$ as an element of the group algebra $\mathbb{R}\mathfrak{M}_k$. Note that the entries of this matrix are interpreted as subprobability measures. Their total mass $\|M_{ij}\|_1$ are the transition probabilities for the Markov chain $\{J_n\}$. A useful consequence of this definition is that for $\mu = \mu_1 * \mu_2$ we have $M = M_1 M_2$.

For an $m \times m$ matrix M , consider its block decomposition

$$M = \left(\begin{array}{c|c} M_{00} & M_{0*} \\ \hline M_{*0} & M_{**} \end{array} \right),$$

so that M_{**} is an $(m-1) \times (m-1)$ matrix.

Proposition 6.1. *Let M be as above, then*

$$T\mu = M_{00} + M_{0*}(I - M_{**})^{-1}M_{*0},$$

where $(I - A)^{-1}$ means $I + A + A^2 + \cdots$.

This appears as Theorem 2.3 in Kaimanovich (2005). Note that M_{**} is an $(m-1) \times (m-1)$ matrix with entries in $\mathbb{R}G$. We include the above discussion and the brief proof because similar ideas are used next for measures on patterns.

Proof. With the above notation, consider the trajectory of 0 until the first time τ such that $J_\tau = 0$. The probability of a particular trajectory $0 = J_0, J_1, \dots, J_\tau = 0$ is a product of transition probabilities $\|M_{ij}\|_1$ of the J Markov chain. Conditioned on these values, the process Y_n is a product of independent samples from the probability measures $M_{ij}/\|M_{ij}\|_1$ corresponding to the transitions. Paths of length τ correspond to the term $M_{0*}M_{**}^{\tau-2}M_{*0}$, and $\tau = 1$ gives the M_{00} term. \square

Ascension formula for patterns

Suppose now that μ is a probability measure on \mathcal{Q}_k , and consider the random walk on \mathfrak{M}_k with step distribution $\bar{\mu}$. This measure has a great deal of symmetry. A uniformly chosen term $X \in \mathfrak{L}_{-1}$ sends any i to a uniform vertex on level 1 of the tree, and has trivial sections $X(i)$. An element $X \in \mathfrak{L}_i$ for $i > 0$ stabilizes the first level of the tree, and its sections are $X(0) = X$ and $X(i) \in \mathfrak{L}_{i-1}$, again uniformly. Similarly, an element $X \in \mathfrak{L}_0$ fixes 0, permutes $\{1, \dots, m-1\}$ uniformly, and its sections are $X(0) = X$ and $X(i) \in \mathfrak{L}_{i-1}$, again uniformly and independently.

Thus the matrix M defined above is as follows:

- For \bar{q}_ℓ , $\ell > 0$ we get a diagonal matrix, with terms $M_{00} = \bar{q}_\ell$ and $M_{ii} = \overline{q_{\ell-1}}$.
- For $\overline{q_{-1}}$ it is the constant matrix $M_{ij} = 1/m$ (entries are the measure of mass $1/m$ on the identity element of the group).
- For \bar{q}_0 we have $M_{00} = \overline{q_{-1}}$, and $M_{0i} = M_{i0} = 0$ for $i \neq 0$. The remaining minor has constant entries: $M_{ij} = \frac{1}{m-1} \overline{q_{-1}}$ for $i, j > 0$.

Note that in all cases, all entries M_{ij} are in $\mathbb{R}\overline{\mathcal{Q}_k}$.

Next, consider a measure \bar{p} for a pattern $p = q_{\alpha_1} \cdots q_{\alpha_\ell}$. Since \bar{p} is a convolution of $\overline{q_{\alpha_i}}$, the resulting M is a product of the corresponding M 's, and also has entries in $\mathbb{R}\overline{\mathcal{Q}_k}$. Finally, from (3) we see that M is linear in the step distribution. Since $\bar{\mu}$ is a convex combination of \bar{p} , we see that $\bar{\mu}$ also gives M with entries in $\mathbb{R}\overline{\mathcal{Q}_k}$.

Further more, note that for all \bar{q}_ℓ we have that the following sets of entries are all constant:

- the first row, except M_{00} ;
- the first column, except M_{00} ;
- the main diagonal, except M_{00} ;
- $M_{i,j}$ for $i \neq j$ and $i, j > 0$.

Matrices with these properties form an algebra, and therefore the same identities hold for any $\bar{\mu}$. This allows us to simplify some subsequent computations by using the 2×2 matrices R defined by

$$\begin{aligned} R_{00} &= M_{00}, & R_{01} &= \sum_{j>0} M_{0j}, \\ R_{10} &= M_{10}, & R_{11} &= \sum_{j>0} M_{1j}. \end{aligned}$$

where M is defined by (3) for a sample of $\bar{\mu}$. (Taking any $i > 0$ in place of 1 will not make a difference.) Since $M \mapsto R$ is a homomorphism of matrix algebras, $\mu \mapsto R$ is a homomorphism from $\mathbb{R}\overline{\mathcal{Q}_k}$ to the matrix algebra of 2×2 matrices over $\mathbb{R}\mathfrak{M}_k$: if $\mu = \mu_1 * \mu_2$ then $R = R_1 R_2$.

We now introduce the operators \mathcal{R}_{ij} , so that for any measure μ on \mathcal{Q}_k we have $R_{ij}\bar{\mu} = \overline{\mathcal{R}_{ij}\mu}$. Since $\mu \mapsto R$ is a homomorphism, it suffices to define \mathcal{R} for patterns of a single letter, and extend it by multiplicativity and linearity to all of $\mathbb{R}\mathcal{Q}_k$. It is clear that the following definition satisfies this desire. For $k \geq 0$, define

$$\mathcal{R}_{ij}q_k = \begin{cases} q_k & i = j = 0, \\ q_{k-1} & i = j = 1, \\ 0 & i \neq j. \end{cases}$$

Note that $k = 0$ no longer needs special care (an advantage of using R over M). For $k = -1$ define

$$\mathcal{R}_{ij}q_{-1} = \begin{cases} \frac{1}{m} & j = 0, \\ \frac{m-1}{m} & j = 1. \end{cases}$$

It follows that the total mass of $\mathcal{R}_{ij}\nu$ for any measure on patterns is given by

$$\mathcal{R}_{ij}\nu(\mathcal{Q}_k) = 1_{i=j}(1 - \nu(D)) + \nu(D)\frac{(m-1)^j}{m}, \quad (4)$$

where D is the set of patterns containing the letter q_{-1} . Let $\mathcal{R}\mu$ to be the matrix with entries $\mathcal{R}_{ij}\mu$. As noted above, for a pattern $p = q_{\alpha_1} \cdots q_{\alpha_\ell}$ we have, using matrix multiplication, $\mathcal{R}p = \prod \mathcal{R}q_{\alpha_i}$. For measures μ we define $\mathcal{R}\mu = \sum_p \mu(p)\mathcal{R}p$.

A key consequence of this construction is that a version of Proposition 6.1 holds with \mathcal{R} . Define the **pattern ascension** operator by

$$\mathcal{T}\mu = \mathcal{R}_{00}\mu + (\mathcal{R}_{01}\mu)(1 - \mathcal{R}_{11}\mu)^{-1}(\mathcal{R}_{10}\mu), \quad (5)$$

where $(1 - a)^{-1} = 1 + a + a^2 + \cdots$.

Proposition 6.2. *For a probability measure $\mu \in \mathbb{R}\mathcal{Q}_k$ we have $\mathcal{T}\bar{\mu} = \overline{\mathcal{T}\mu}$.*

Proof. Let M be the matrix corresponding to $\bar{\mu}$. By Proposition 6.1,

$$\mathcal{T}\bar{\mu} = M_{00} + M_{0*}(I - M_{**})^{-1}M_{*0}.$$

However, since M is in the aforementioned matrix algebra, it is straightforward to verify that $R_{01}R_{11}^k R_{10} = M_{0*}M_{**}^k M_{*0}$, and hence

$$\mathcal{T}\bar{\mu} = R_{00} + R_{01}(I - R_{11})^{-1}R_{10}.$$

Since $R\bar{\mu} = \overline{\mathcal{R}\mu}$, this implies the proposition. \square

We now inspect the structure of \mathcal{R} in more detail. Since \mathcal{R} is multiplicative, for some patterns its entries are delta measures:

Lemma 6.3. For a pattern $p = q_{\alpha_1} \cdots q_{\alpha_\ell}$ that does not contain the letter q_{-1} , we have

$$\mathcal{R}_{ij}p = \begin{cases} p & i = j = 0, \\ q_{\alpha_1-1} \cdots q_{\alpha_\ell-1} & i = j = 1, \\ 0 & i \neq j. \end{cases}$$

For a general pattern with decomposition $p = p_0 q_{-1} p_1 \cdots q_{-1} p_\ell$ into patterns $p_i \in D^c$ we have

$$\mathcal{R}_{ij}p = (\mathcal{R}_{ii}p_0) \prod_{i=1}^{\ell-1} \left(\frac{1}{m} \mathcal{R}_{00}p_i + \frac{m-1}{m} \mathcal{R}_{11}p_i \right) (\mathcal{R}_{jj}p_\ell).$$

Proof. Since \mathcal{R} is multiplicative, the first part holds by definition. The second part follows from the first and from $\mathcal{R}_{ij}q_{-1} = (m-1)^j/m$. \square

Define the **averaged legacy operator**

$$\mathcal{A} = \sum_{i,j \in \{0,1\}} \frac{(m-1)^i}{m} \mathcal{R}_{ij},$$

so that $\mathcal{A}\mu = \begin{pmatrix} \frac{1}{m} & \frac{m-1}{m} \\ & 1 \end{pmatrix} (\mathcal{R}\mu) \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Note that when μ is a probability measure, so are $\sum_j \mathcal{R}_{ij}\mu$, and $\mathcal{A}\mu$.

Lemma 6.4. Let $p = p_0 q_{-1} p_1 q_{-1} \cdots q_{-1} p_k$ where $p_i \in D^c$ be the decomposition of p into subpatterns not containing q_{-1} (possibly p_0 or p_k are \emptyset). Then

$$\mathcal{A}p = (\mathcal{A}p_0) \cdots (\mathcal{A}p_\ell). \quad (6)$$

Proof. This follows from the fact that \mathcal{R} is multiplicative. Writing $\mathcal{R}q_{-1} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} \frac{1}{m} & \frac{m-1}{m} \end{pmatrix}$ we have

$$\begin{aligned} \prod_{i=0}^{\ell} \mathcal{A}p_i &= \begin{pmatrix} \frac{1}{m} & \frac{m-1}{m} \\ & 1 \end{pmatrix} (\mathcal{R}p_0) \begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} \frac{1}{m} & \frac{m-1}{m} \\ & 1 \end{pmatrix} (\mathcal{R}p_1) \begin{pmatrix} 1 \\ 1 \end{pmatrix} \cdots \begin{pmatrix} \frac{1}{m} & \frac{m-1}{m} \\ & 1 \end{pmatrix} (\mathcal{R}p_\ell) \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{m} & \frac{m-1}{m} \\ & 1 \end{pmatrix} (\mathcal{R}p_0) (\mathcal{R}q_{-1}) (\mathcal{R}p_1) \cdots (\mathcal{R}q_{-1}) (\mathcal{R}p_\ell) \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= \mathcal{A}p. \end{aligned} \quad \square$$

Algorithmic description of \mathcal{T}

We now present an algorithm that uses samples of μ to get a sample from $\mathcal{T}\mu$. This algorithm is an interpretation of Proposition 6.2, though it is also possible to use it to define \mathcal{T} , and derive Proposition 6.2 from it analogously to Proposition 6.1.

To sample $\mathcal{T}\mu$ we consider the Markov chain $\{\mathcal{E}_n, Q_n\}_{n>0}$ with state space $\{0, 1\} \times \mathcal{Q}_k$ and transition probabilities

$$\mathbb{P}(\mathcal{E}_n = j, Q_n = p | \mathcal{E}_{n-1} = i, Q_{n-1}) = (\mathcal{R}_{ij}\mu)(p). \quad (7)$$

Lemma 6.5. *If $\mathcal{E}_0 = 0$ and $\tau > 0$ is minimal such that $\mathcal{E}_\tau = 0$, then $Q_1 \cdots Q_\tau$ has the law of $\mathcal{T}\mu$.*

Proof. This is almost identical to Proposition 6.1. The term $\mathcal{R}_{00}\mu$ corresponds to $\tau = 1$, and the term $(\mathcal{R}_{01}\mu)(\mathcal{R}_{11}\mu)^n(\mathcal{R}_{10}\mu)$ to $\tau = n + 2$. \square

Since $\mathcal{R}\mu$ is linear, the algorithm above can be broken into to the following steps

- Start with $\mathcal{E}_0 = 0$.
- For $i = 1, 2, \dots$ sequentially pick $P_n \sim \mu$, and pick (\mathcal{E}_n, Q_n) with distribution

$$\mathbb{P}(\mathcal{E}_n = j, Q_n = p | \mathcal{E}_{n-1} = i, Q_{n-1}) = (\mathcal{R}_{ij}P_n)(p).$$

- At the first time $\tau \geq 1$ so that $\mathcal{E}_\tau = 0$, stop and return the pattern $Q_1 \cdots Q_\tau$.

7 Properties of \mathcal{T} and the sequence $\mathcal{T}^k\mu$

In this section we study the evolution of a measure μ on patterns under iterated ascension.

Proposition 7.1. *For a probability measure μ on patterns, the moments of the pattern lengths with respect to $\mathcal{T}\mu$ satisfy*

$$\mathcal{T}\mu(|p|) \leq m\mathcal{A}\mu(|p|), \quad \mathcal{T}\mu(|p|^2) \leq \frac{2m^3}{(m-1)\mu(D)}\mathcal{A}\mu(|p|)^2 + m\mathcal{A}\mu(|p|^2).$$

Proof. Because length is subadditive ($|pq| \leq |p| + |q|$), (5) implies that the first and second moment of the length of a sample from $\mathcal{T}\mu$ are bounded above by the first and second moment of

$$\theta = \tilde{\mathcal{R}}_{00}\mu + \tilde{\mathcal{R}}_{01}\mu \left(1 - \tilde{\mathcal{R}}_{11}\mu\right)^{-1} \tilde{\mathcal{R}}_{10}\mu,$$

where $\tilde{\mathcal{R}}_{ij}$ refer to induced measures on length, and the above formula takes values in the semi-group algebra of \mathbb{Z}_+ . Let $\varphi_{ij}(z) = \sum_n (\tilde{\mathcal{R}}_{ij}\mu)(n)z^n$ denote the generating functions for the measures $\tilde{\mathcal{R}}_{ij}\mu$. The generating function of the measure θ is

$$f = \varphi_{00} + \varphi_{01}(1 - \varphi_{11})^{-1}\varphi_{10}.$$

The first and second moments of θ are given by $f'(1)$ and $f'(1) + f''(1)$ respectively.

We have formula (4) for the total mass:

$$\varphi_{ij}(1) = \mathcal{R}_{ij}\mu(\mathcal{Q}_k) = 1_{i=j}(1 - \mu(D)) + \mu(D)\frac{(m-1)^j}{m}.$$

It follows that

$$f'(1) = \varphi'_{00}(1) + \varphi'_{01}(1) + (m-1)(\varphi'_{10}(1) + \varphi'_{11}(1)) = m\mathcal{A}\mu(|p|).$$

The second moment is computed in a similar fashion, using second derivatives. With the notation $\kappa_{ij} = \varphi'_{ij}(1) = R_{ij}\mu(|p|)$, a straightforward computation gives the second moment $f''(1) + f'(1)$ of θ as

$$\frac{2m}{\mu(D)}(\kappa_{10} + \kappa_{11})(\kappa_{01} + (m-1)\kappa_{11}) + m\mathcal{A}\mu(|p|^2) \leq \frac{2m}{\mu(D)}\frac{(m\mathcal{A}\mu(|p|))^2}{m-1} + m\mathcal{A}\mu(|p|^2). \quad \square$$

Lemma 7.2. *We have the following (where $D \subset \mathcal{Q}_k$ is the set of patterns containing q_{-1}):*

- (a) *For any $x \notin D$ we have $\mathcal{T}\mu(x) \geq \mu(x)$.*
- (b) *If $\mu(D) > 0$, then $\mathcal{T}\mu(D) < \mu(D)$.*
- (c) *If μ is supported on $\mathcal{Q}_{k-1} \cup \{q_k\}$ then so is $\mathcal{T}\mu$, and $\mu(q_k) = \mathcal{T}\mu(q_k)$.*

Proof. We refer to the algorithmic description of $\mathcal{T}\mu$ and Lemma 6.3. If the first pattern P_1 chosen from μ is in D^c , then $\mathcal{E}_1 = 0$ and $Q_1 = P_1$ is also the output of the procedure. Claim (a) follows.

For part (b), assume that $\mu(p) > 0$ for some $p \in D$. Consider $\mathcal{R}p$ as given by Lemma 6.3. Taking only the $\frac{1}{m}\mathcal{R}_{00}$ terms in the product shows that $\mathcal{R}_{00}p$ assigns some positive probability $\delta \geq m^{-\ell}$ to the pattern p with all appearances of q_{-1} deleted. This pattern is in D^c . Thus $\mathcal{T}\mu(D^c) \geq \mu(D^c) + \delta\mu(p)$.

For part (c), note that for such measures, the only way to get q_k in the ascension algorithm is if the first pattern selected from μ is q_k , in which case it is also the output. \square

Lemma 7.3. *If the measure sequence $\{\mathcal{T}^k\mu\}$ is tight, then its limit exists and is supported on D^c .*

Proof. Since the measures are tight they have subsequential limits that are probability measures. Let ν be such a limit. \mathcal{T} is continuous, so there are k so that ν and $\mathcal{T}^k\mu$ are close enough to have

$$\mathcal{T}^{k+1}\mu(D) - \mathcal{T}\nu(D) < \varepsilon.$$

But again since ν is a limit point there is $\ell > k$ so that

$$\nu(D) - \mathcal{T}^\ell\mu(D) < \varepsilon.$$

Summing the last two formulae and using that $\mathcal{T}^\ell \mu(D) \leq \mathcal{T}^{k+1} \mu(D)$ we get

$$\nu(D) - \mathcal{T}\nu(D) < 2\varepsilon.$$

Since ε is arbitrary, $\mathcal{T}\nu(D) = \nu(D)$. By Lemma 7.2(b), this implies that $\nu(D) = 0$.

Now, $\mathcal{T}^k \mu(x)$ is monotone in k for $x \notin D$, so all subsequential limits are equal. \square

Lemma 7.4. *Consider an entropy-tight set of measures $\{\mu_i\}$ on \mathcal{Q}_k , such that $\mu_i(|p|) < M$ for some M . Assume further that for patterns $p \in \bigcup_i \text{supp}(\mu_i)$ we have $H(\bar{p}) = o(|p|)$. Then $\{\bar{\mu}_i\}$ is also entropy-tight.*

Proof. Denote $K = \{g \in \mathfrak{M}_{m,k} : |g| \leq \ell\}$, and consider a measure μ on \mathcal{Q}_k . By definition,

$$\bar{\mu}(\cdot) = \sum_p \mu(p) \bar{p}(\cdot).$$

For compactness, denote $h(x) = -x \log x$, and note that h is subadditive. We have

$$H_{K^c}(\bar{\mu}) = \sum_{|g|>\ell} h(\bar{\mu}(g)) \leq \sum_{|g|>\ell} \sum_p h(\mu(p) \bar{p}(g)).$$

Since \bar{p} is supported on group elements of length at most $|p|$, the above equals

$$\sum_{|g|>\ell} \sum_{|p|>\ell} h(\mu(p) \bar{p}(g)) \leq \sum_{|p|>\ell} \sum_g h(\mu(p) \bar{p}(g)).$$

However, since \bar{p} is a probability measure, for any $a > 0$ we have

$$\sum_g h(a \bar{p}(g)) = - \sum_g (\log a + \log \bar{p}(g)) a \bar{p}(g) = h(a) + a H(\bar{p}),$$

and therefore

$$H_{K^c}(\bar{\mu}) \leq \sum_{|p|>\ell} h(\mu(p)) + \sum_{|p|>\ell} \mu(p) H(\bar{p}). \quad (8)$$

Now fix ε . For ℓ sufficiently large, the first sum in (8) is uniformly small for all μ in our entropy-tight family. Also by assumption, if ℓ is large enough then either $\mu(p) = 0$ or $H(\bar{p}) < \varepsilon|p|$. Thus for large enough ℓ we have

$$H_{K^c}(\bar{\mu}) \leq \varepsilon + \sum_{|p|>\ell} \mu(p) \varepsilon |p| \leq \varepsilon + M\varepsilon.$$

Since ε is arbitrary, this implies that $\{\bar{\mu}_i\}$ is entropy-tight as claimed. \square

8 Preliminary results for entropy of patterns

Here we establish some useful facts about entropy of $\bar{\mu}$ for some measures μ on patterns. We first recover the following result.

Theorem 8.1 (Bartholdi, Kaimanovich, and Nekrashevych (2008)). *Let μ be supported on patterns of length 1 in \mathcal{Q}_0 . Then $H_\infty(\bar{\mu}) = 0$, or equivalently, the entropy $H(\bar{\mu}^{*n}) = o(n)$.*

We present what is essentially the original proof, written in the language of this paper.

Proof. Lemma 7.2(c) shows that $\mathcal{T}^k\mu$ is supported on $\{q_0\} \cup \mathcal{Q}_{-1} = \{q_0, q_{-1}, \emptyset\}$. Thus the sequence $\mathcal{T}^k\bar{\mu}$ is tight and entropy-tight. Lemma 7.3 shows that the $\nu = \lim \mathcal{T}^k\mu$ exists and is supported on $\{q_0, \emptyset\}$. Thus $\text{supp}(\bar{\nu})$ is contained in a finite group, and $H_\infty(\bar{\nu}) = 0$. The result now follows by Theorem 4.2. \square

Next, we show that evaluation of patterns on \mathcal{Q}_0 have entropy which is sublinear in their length. We need the following simple lemma.

Lemma 8.2 (Monotonicity of pattern entropy). *If p and r are patterns, then*

$$H(\overline{pr}) \geq \max(H(\bar{p}), H(\bar{r})).$$

Proof. \overline{pr} is a convex combination of measures of the form $\bar{p}g$ where g is chosen according to \bar{Q} . All these measures have the same entropy $H(\bar{r})$. Entropy is a concave function, so by Jensen's inequality

$$H(\overline{pr}) \geq H(\bar{p}).$$

In the same way, $H(\overline{pr}) \geq H(\bar{r})$. \square

Lemma 8.3. *There is a function $f(n) = o(n)$ such that for any pattern $p \in \mathcal{Q}_0$*

$$H(\bar{p}) \leq f(|p|).$$

Proof. Let μ be the uniform measure on $\{q_{-1}, q_0\}$, and $\mu_n = \mu^{*n}$. By concavity of entropy and Jensen's inequality, since $\bar{\mu}_n = \sum \mu_n(p)\bar{p}$, we have

$$H(\bar{\mu}_n) \geq \sum_{p \in \mathcal{Q}_0} \mu_n(p)H(\bar{p}) \geq \sum_{p \in A_\ell} \mu_n(p)H(\bar{p})$$

where A_ℓ consists of all patterns of length greater than ℓ . By Lemma 8.2, the entropy of each word of length n is greater than the entropy of each word of length $n - 1$ (since patterns in \mathcal{Q}_0 are just alternating q_0 's and q_{-1} 's). It follows that

$$H(\bar{\mu}_n) \geq \mu_n(A_\ell)H(\overline{Q_\ell^*}),$$

where Q_ℓ^* can be either of the two patterns of length ℓ .

The length of a sample of μ_n is the number of runs in the word $Q_1 \cdots Q_n$ with i.i.d. letters. Since each letter starts a new run with probability $1/2$, the length is binomial, and symmetric about $n/2$. Thus with $n = 2\ell$ we have $\mu_n(A_\ell) \geq 1/2$. We find

$$H(Q_\ell^*) \leq 2H(\overline{\mu_{2\ell}}) = o(\ell). \quad \square$$

We can also make some conclusions about linear automaton groups. It is not hard to see that the combined supports $\mathfrak{L}_0, \mathfrak{L}_1$ of the evaluated patterns $\overline{q}_0, \overline{q}_1$ generate a bounded automaton group. Certain walks on this group also have zero asymptotic entropy.

Lemma 8.4. *Any measure μ on \mathcal{Q}_1 supported on the patterns $\{q_0, q_1, \emptyset\}$ satisfies $H_\infty(\overline{\mu}) = 0$.*

Proof. Consider the subgroup of $\mathfrak{M}_{m,1}$ consisting of automorphisms of the form

$$g = \langle\langle g, g_1, \dots, g_{m-1} \rangle\rangle \rho,$$

where $0.\rho = 0$. Note that the support of $\overline{\mu}$ is contained in this subgroup. Consider a random walk X_k with step distribution $\overline{\mu}$, and write

$$X_k = \langle\langle X_k, Y_k(1), \dots, Y_k(m-1) \rangle\rangle \rho_k.$$

The key observation is that the distribution of each $Y_1(i)$ is $\overline{\mu'}$ where $\mu' = \mathcal{R}_{11}\mu$. It follows that for each $i > 0$ the process $(Y_k(i), k \geq 0)$ is a random walk on $\mathfrak{M}_{m,0}$ with step distribution $\overline{\mu'}$, where μ' is supported on $\{q_{-1}, q_0, \emptyset\}$. By Theorem 8.1, this random walk has zero asymptotic entropy. On the other hand, X is determined by ρ and the Y_i 's, so

$$H(X_k) \leq H(\rho_k) + \sum_{i=1}^{m-1} H(Y_k(i)).$$

Thus $H_\infty(\overline{\mu}) \leq (m-1)H_\infty(\overline{\mu'}) = 0$, as required. \square

9 The linear-activity case and the main theorem

The key step in the proof of the main theorem is the following proposition. It is based on results from the previous sections.

Proposition 9.1. *Let μ be a measure on $\mathcal{Q}_0 \cup \{q_1\}$ with $\mu(|p|^2) < \infty$. Then the sequence $\mu_k = \mathcal{T}^k \mu$ is both tight and entropy-tight.*

The tightness of the sequence depends on a certain contracting property of the ascension operator \mathcal{T} for patterns on \mathcal{Q}_0 . Proposition 7.1 gives bounds on moments of $\mathcal{T}\mu$ in terms of moments of $\mathcal{A}\mu$. The next lemma bounds the moments of $\mathcal{A}\mu$.

Lemma 9.2 (Length of legacy). *For a probability measure μ on $\{q_1\} \cup \mathcal{Q}_0$, we have*

$$\mathcal{A}\mu(|p|) \leq 1 + \frac{m-1}{m^2}\mu(|p|), \quad \mathcal{A}\mu(|p|^2) \leq \frac{(m-1)^2}{m^4}\mu(|p|^2) + 3\mu(|p|).$$

Proof. μ is a convex combination of delta measures on a single patterns. Since $\mathcal{A}\mu$ is linear, and the bounds are all affine in μ , it suffices to prove the claim for delta measures $\mu = p$ supported on a single pattern $p \in \{q_1\} \cup \mathcal{Q}_0$.

If $p \in \{\emptyset, q_1, q_{-1}\}$, then $\mathcal{A}\mu$ is supported on patterns of length at most 1, and the claim is trivial.

Otherwise, $\mu = p$ for some $p \in \mathcal{Q}_0$. Let $t > 0$ be the number of times q_0 appears in p , so that $|p| \geq 2t - 1$. By (6), a sample of $\mathcal{A}\mu$ is given by $X_1 \cdots X_t$ where X_i are i.i.d. with distribution $\mathcal{A}q_0 = \frac{1}{m}q_0 + \frac{m-1}{m}q_{-1}$. Each run of q_0 's or q_{-1} 's reduces to a single letter, so the length R of $X_1 \cdots X_t$ is given by the number of runs.

Let N_i be the indicator of the event that X_i starts a new run, so that $R = N_1 + \dots + N_j$. Then $N_1 \equiv 1$ and for $1 < i \leq t$ we have $\mathbb{E}N_i = \frac{2(m-1)}{m^2}$. Now,

$$\mathbb{E}R = 1 + \frac{2(m-1)}{m^2}(t-1) \leq 1 + \frac{m-1}{m^2}|p|.$$

For the second moment, since N_i is independent of N_j unless $|i-j| \leq 1$, for $t \geq 1$ we get

$$\mathbb{E}R^2 = \sum_{i,j \leq t} \mathbb{E}N_i N_j \leq \frac{4(m-1)^2}{m^4}(t-1)^2 + 5t - 6 \leq \frac{(m-1)^2}{m^4}|p|^2 + 3|p|.$$

Here, $5t - 6$ is the number of pairs that are dependent or involve N_1 . The right hand side is a bound for the $\mathbb{E}(R^2|J)$ even for $J = 0$. \square

Combining Lemma 9.2 with Proposition 7.1 gives the following contraction property of \mathcal{T} :

Corollary 9.3 (Contraction). *For any probability measure μ on $\mathcal{Q}_0 \cup \{q_1\}$ we have*

$$\mathcal{T}\mu(|p|) \leq \frac{m-1}{m}\mu(|p|) + m, \quad \mathcal{T}\mu(|p|^2) \leq \frac{2}{\mu(D)}(\mu(|p|) + m)^2 + \frac{(m-1)^2}{m^3}\mu(|p|^2).$$

Proof of Proposition 9.1. Any non-negative sequence satisfying $x_{k+1} \leq c + \alpha x_k$ with $\alpha \in [0, 1)$ must be bounded. By the first claim of Lemma 9.3, the sequence $\mu_k(|p|)$ is bounded, which implies tightness.

Similarly, if $\varepsilon_k \geq 0$ is a non-increasing sequence, then for any sequence $x_k \geq 0$ the inequalities $x_{k+1} \leq c/\varepsilon_k + \alpha x_k$ imply that $\varepsilon_k x_k$ is bounded. Set $\varepsilon_k = \mu_k(D)$, which is non-increasing by Lemma 7.2(b). With the second claim of Lemma 9.3, and $\mu_k(|p|) < c$, this implies that the sequence $\mu_k(D) \cdot \mu_k(|p|^2)$ is also bounded.

For entropy-tightness, Since μ_k is supported within $\{q_0, q_1, \emptyset\} \cup D$, it suffices to show that the contribution to its entropy from the set D converges to 0. Let ν_k denote μ_k conditioned to the set D . The contribution to $H(\mu_k)$ from D is given by

$$\sum_{P \in D} -\mu_k(P) \log \mu_k(P) = \varepsilon_k H(\nu_k) - \varepsilon_k \log \varepsilon_k. \quad (9)$$

We have

$$\nu_k(|p|^2) \leq \frac{\mu_k(|p|^2)}{\varepsilon_k} \leq \frac{C}{\varepsilon_k^2}.$$

However, note that ν_k is supported on patterns of alternating q_0 's and q_{-1} 's, so we have

$$H(\nu_k) \leq 1 + H(\tilde{\nu}_k) \quad (10)$$

where $\tilde{\nu}_k$ is the induced measure on lengths. Lemma 3.1 bounds entropy in terms of variance, and hence in terms of second moment. In our case it gives

$$H(\tilde{\nu}_k) \leq 1 + C + (\log(C/\varepsilon_k^2))^+ \leq C' + |\log \varepsilon_k^2|.$$

Combining this with (9), (10), and Lemma 7.3 which shows that $\varepsilon_k \rightarrow 0$, completes the proof. \square

We are ready to prove the following, which implies Theorem 1.

Theorem 9.4. *Let μ be a probability measure on patterns in $\mathcal{Q}_0 \cup \{q_1\}$ with $\mu(|p|^2) < \infty$. The random walk on the linear-activity mother group with step distribution $\bar{\mu}$ has zero asymptotic entropy.*

Proof. By Theorem 4.2 it suffices to show that the sequence $T^k \bar{\mu}$ is entropy-tight and converges weakly to a measure with zero asymptotic entropy.

Consider first the sequence $\mathcal{T}^k \mu$ of measures on patterns. Proposition 9.1 shows this sequence is tight. Lemmas 7.2 and 7.3 imply that $\nu = \lim \mathcal{T}^k \mu$ exists and is supported on $\{q_1, q_0, \emptyset\}$. Finally, by Lemma 8.4, evaluation of measures supported on $\{q_1, q_0, \emptyset\}$ have zero entropy, so $H_\infty(\bar{\nu}) = 0$.

It remains to show that $T^k \bar{\mu}$ is entropy-tight. This follows from Lemma 7.4, so we verify its conditions. Entropy-tightness and uniformly bounded expected length for $\mathcal{T}^k \mu$ are proved in Proposition 9.1. The support of any $\mathcal{T}^k \mu$ is contained in $\{q_1\} \cup \mathcal{Q}_0$. By Lemma 8.3 the entropy of $p \in \{q_1\} \cup \mathcal{Q}_0$ is $o(|p|)$. \square

Acknowledgements. We thank V. Nekrashevich for bringing the example of the long-range group to our attention.

References

- L. Bartholdi, V. A. Kaimanovich, and V. V. Nekrashevych. On amenability of automata groups, 2008. [arXiv.org:0802.2837](https://arxiv.org/abs/0802.2837).
- L. Bartholdi and B. Virág (2005). Amenability via random walks. *Duke Math. J.*, 130(1):39–56.
- I. Benjamini and C. Hoffman (2005). ω -periodic graphs. *Electron. J. Combin.*, 12:Research Paper 46, 12 pp. (electronic).
- Y. Glasner and S. Mozes (2005). Automata and square complexes. *Geom. Dedicata*, 111:43–64.
- R. Grigorchuk and Z. Šunić (2006). Asymptotic aspects of Schreier graphs and Hanoi Towers groups. *C. R. Math. Acad. Sci. Paris*, 342(8):545–550.
- R. I. Grigorchuk (1984). Degrees of growth of finitely generated groups and the theory of invariant means. *Izv. Akad. Nauk SSSR Ser. Mat.*, 48(5):939–985.
- R. I. Grigorchuk and A. Żuk (2001). The lamplighter group as a group generated by a 2-state automaton, and its spectrum *Geom. Dedicata*, 87(1-3):209-244.
- R. I. Grigorchuk and A. Żuk (2002). On a torsion-free weakly branch group defined by a three state automaton. *Internat. J. Algebra Comput.*, 12(1-2):223–246.
- Guido’s book of conjectures* (2008). Collected by I. Chatterji. *Enseign. Math. (2)*, 54(1-2): 3–189.
- V. A. Kaimanovich (2005). “Münchhausen trick” and amenability of self-similar groups. *Internat. J. Algebra Comput.*, 15(5-6):907–937.
- A. W. Kemp (1997). Characterizations of a discrete normal distribution. *J. Statist. Plann. Inference*, 63(2):223–229.
- The Kourovka notebook*. V. D. Mazurov and E. I. Khukhro, editors. Russian Academy of Sciences, Novosibirsk, sixteenth edition, 2006. Unsolved problems in group theory.
- J. Milnor (1968). Growth of finitely generated solvable groups. *J. Differential Geometry*, 2: 447–449.
- S. Sidki (2000). Automorphisms of one-rooted trees: growth, circuit structure, and acyclicity. *J. Math. Sci. (New York)*, 100(1):1925–1943. Algebra, 12.
- Z. Sunic and R. I. Grigorchuk (2007). Self-similarity and branching in group theory. *London Mathematical Society Lecture Note Series* 339, 36–95.

M. Vorobets and Y. Vorobets (2007). On a free group of transformations defined by an automaton. *Geom. Dedicata*, 124:237–249.