ASYMPTOTIC ENTROPY OF RANDOM WALKS ON REGULAR LANGUAGES OVER A FINITE ALPHABET

LORENZ A. GILCH

ABSTRACT. We prove existence of asymptotic entropy of random walks on regular languages over a finite alphabet and we give formulas for it. Furthermore, we show that the entropy varies real-analytically in terms of the probability measures of constant support, which describe the random walk. This setting applies, in particular, to random walks on virtually free groups.

1. INTRODUCTION

Let $\mathcal{A}$ be a finite alphabet and let $\mathcal{A}^*$ be the set of all finite words over the alphabet $\mathcal{A}$, where o denotes the empty word. Consider a transient Markov chain $(X_n)_{n \in \mathbb{N}_0}$ on $\mathcal{A}^*$ with $X_0 = o$ such that the transition probabilities depend only on the last $K \in \mathbb{N}$ letters of the current word, in between two steps the word length differs only by at most K letters and in each step only the last $K \in \mathbb{N}$ letters of the current word may be modified. Denote by π_n the distribution of X_n. We are interested whether the sequence $\frac{1}{n}\mathbb{E}[-\log \pi_n(X_n)]$ converges, and if so to describe the limit. If it exists, it is called the *asymptotic entropy*, which was introduced by Avez [1]. The aim of this paper is to prove existence of the asymptotic entropy, to describe it as the rate of escape w.r.t. the Greenian distance and to prove its real-analytic behaviour when varying the transition probabilities of constant support.

We outline some background on this topic. It is well-known by Kingman's subadditive ergodic theorem (see Kingman [11]) that the entropy exists for random walks on groups if $\mathbb{E}[-\log \pi_1(X_1)] < \infty$. In contrast to this fact existence of the entropy on general structures is not known a priori. In our setting we are not able to apply the subadditive ergodic theorem since we neither have subadditivity nor a global composition law of words if we restrict the random walk to be on a proper subset of $\mathcal{A}^*$. This forces us to use other techniques like generating functions techniques. These generating functions are power series with probabilities as coefficients, which describe the characteristic behaviour of the underlying random walks. The technique of our proof of existence of the entropy was motivated by Benjamini and Peres [2], where it is shown that for random walks on groups the entropy equals the rate of escape w.r.t. the Greenian distance; compare also with Blachère, Haïssinsky and Mathieu [3]. In particular, we will also show that the asymptotic entropy h is the rate of

escape w.r.t. a distance function in terms of Green functions, which in turn yields that h is also the rate of escape w.r.t. the Greenian distance. Moreover, we prove convergence in probability and convergence in L_1 of the sequence $-\frac{1}{n}\log \pi_n(X_n)$ to h, and we show also that h can be computed along almost every sample path as the limes inferior of the afore-mentioned sequence. The question of almost sure convergence of $-\frac{1}{n}\log \pi_n(X_n)$ to some constant h, however, remains open. Similar results concerning existence and formulas for the entropy are proved in Gilch and Müller [8] for random walks on directed covers of graphs and in Gilch [7] for random walks on free products of graphs. Furthermore, we give formulas for the entropy which allow numerical computations and also exact calculations in some special cases.

Kaimanovich and Erschler asked whether drift and entropy of random walks vary continuously (or even analytically) when varying the probabilities of the random walk with keeping the support of single step transitions constant. In this article we also show that h is real-analytic in terms of the parameters describing the random walk on $\mathcal{A}^*$. This fact applies, in particular, to the case of bounded range random walks on virtually free groups. At this point let us mention that several papers concerning continuity and analyticity of the drift and entropy have been published recently: e.g., see Ledrappier [13], [14], Haïssinsky, Mathieu and Müller [9], Gilch [7]. The recent article of Gilch and Ledrappier [5] collects several results about analyticity of drift and entropy of random walks on groups.

The reasoning of our proofs follows a similar argumentation as in [8] and [7]: we will show that the entropy equals the rate of escape w.r.t. some special length function, and we deduce the proposed properties analogously. The plan of the paper is as follows: in Sections 2 and 3 we define the random walk on the regular language and the associated generating functions. Sections 4 explains the structure of cones in the present context. In Sections 5 and 6 we prove existence of the asymptotic entropy, while in Section 7 we give explicit formulas for it. Section 8 shows real-analyticity of the entropy.

2. Notation

Let $\mathcal{A}$ be a finite alphabet and denote by o the empty word. A *random walk on a regular language* is a Markov chain on a subset $\mathcal{L} \subseteq \mathcal{A}^* := \bigcup_{n\geq 1} \mathcal{A}^n \cup \{o\}$ of all finite words over the alphabet $\mathcal{A}$, whose transition probabilities obey the following rules:

  (i) Only the last two letters of the current word may be modified.
 (ii) Only one letter may be adjoined or deleted at one instant of time.
(iii) Adjunction and deletion may only be done at the end of the current word.
(iv) Probabilities of modification, adjunction or deletion depend only on the last two letters of the current word.

Compare with Lalley [12] and Gilch [6]. The assumption that transition probabilities depend only on the last two letters of the current word may be weakened to dependence of the last $K \geq 2$ letters by blocking words of length at most K to new letters (compare with [12, Section 3.3]. In general, a regular language is a subset of $\mathcal{A}^*$ whose words are accepted by a finite-state automaton. It is necessary that by each modification of a word of the regular language in one single step a new word of the regular language is created. The

results below, however, are so general such that w.l.o.g. – for ease and better readability – we may assume that the regular language $\mathcal{L}$ consists of the whole set $\mathcal{A}^*$. We will use the notation $w \in \mathcal{L}$, $w \in \mathcal{A}^*$ respectively, to emphasize at some points that we explicitely mean a word of the language or just a word over the alphabet. Let us note that random walks on *virtually free groups* constitute a special case of our setting, and our results directly apply.

We introduce some notation. For a word $w \in \mathcal{L}$ and $k \in \mathbb{N}$, $w[k]$ denotes the k-th letter of w, and $[w]$ denotes the last two letters of w. The random walk on $\mathcal{L}$ is described by the sequence of random variables $(X_n)_{n \in \mathbb{N}_0}$. Initially, we have $X_0 := o$. If we want to start the random walk at $w \in \mathcal{L}$ instead of o, we write for short $\mathbb{P}_w[\cdot] := \mathbb{P}[\cdot \mid X_0 = w]$. For two words $w_1, w_2 \in \mathcal{A}^*$, we write $w_1 w_2$ for the concatenated word. We use the following abbreviations for the transition probabilities: for $w \in \mathcal{L}$, $a_1, a_2, b_1 \in \mathcal{A}$, $b_2, c \in \mathcal{A} \cup \{o\}$, $n \in \mathbb{N}_0$, we write

$$\mathbb{P}[X_{n+1} = wa_2b_2c \mid X_n = wa_1b_1] = p(a_1b_1, a_2b_2c),$$
$$\mathbb{P}[X_{n+1} = b_2c \mid X_n = a_1] = p(a_1, b_2c),$$
$$\mathbb{P}[X_{n+1} = b_2 \mid X_n = o] = p(o, b_2).$$

We assume that $\mathbb{P}_u[X_1 = v] > 0$ implies $\mathbb{P}_v[X_1 = u] > 0$ for all $u, v \in \mathcal{L}$. We call this property *weak symmetry*. For $w_1, w_2 \in \mathcal{L}$, the n-step transition probabilities are denoted by $p^{(n)}(w_1, w_2) := \mathbb{P}_{w_1}[X_n = w_2]$. The *natural word length* of any $w \in \mathcal{L}$ is denoted by $|w|$. Malyshev [15] proved that the rate of escape w.r.t. the natural word length exists under some natural assumptions, that is, there is a non-negative constant ℓ such that

$$\lim_{n \to \infty} \frac{|X_n|}{n} = \ell \quad \text{almost surely.}$$

Here, ℓ is called the *rate of escape*. Furthermore, by [15] follows that ℓ is strictly positive if and only if $(X_n)_{n \in \mathbb{N}_0}$ is transient. In [6] there are explicit formulas for the rate of escape w.r.t. more general length functions.

Another characteristic number of random walks is the asymptotic entropy. Denote by π_n the distribution of X_n. If there is a non-negative constant h such that the limit

$$h = \lim_{n \to \infty} -\frac{1}{n} \mathbb{E} \log \pi_n(X_n)$$

exists, then h is called the *asymptotic entropy*. Since we only have a partial composition law for concatenation of two words (if $\mathcal{L} \subset \mathcal{A}^*$) and since we have no subadditivity and transitivity of the random walk, we can not apply – as in the case of random walks on groups – Kingman's subadditive ergodic theorem to show existence of h. It is easy to see that the entropy equals zero if the random walk is recurrent (see Corollary 7.4). Therefore, we assume from now on transience of $(X_n)_{n \in \mathbb{N}_0}$.

Moreover, we assume that the random walk on $\mathcal{L}$ is *suffix-irreducible*, that is, for all $w \in \mathcal{L}$ with $\mathbb{P}[X_m = w] > 0$ for some $m \in \mathbb{N}$ and for all $ab \in \mathcal{A}^2$ there is some $n \in \mathbb{N}$ such that

$$\mathbb{P}\Big[\exists w_1 \in \mathcal{A}^* : X_n = ww_1ab, \forall k < n : |X_k| \geq |w| \;\Big|\; X_0 = w\Big] > 0.$$

This assumption excludes degenerate cases and guarantees existence of ℓ; compare with [6, End of Section 2.1]. At this point let us mention that $\lim_{n \to \infty} -\frac{1}{n} \log \pi_n(X_n)$ is not necessarily deterministic: take two homogeneous trees of different degrees $d_1, d_2 \geq 3$ equipped

with simple random walk; identify their root with one single root which becomes o; then the limit depends on the fact in which of the two trees the random walks goes to infinity.

3. Generating Functions

For $w_1, w_2 \in \mathcal{L}$, $z \in \mathbb{C}$, the *Green function* is defined as

$$G(w_1, w_2|z) := \sum_{n \geq 0} p^{(n)}(w_1, w_2) \cdot z^n,$$

the *last visit generating function* as

$$L(w_1, w_2|z) := \sum_{n \geq 0} \mathbb{P}\big[X_n = w_2, \forall m \in \{1, \ldots, n\} : X_m \neq w_1 \big| X_0 = w_1\big] \cdot z^n$$

and the *first return generating function* as

$$U(w_1, w_1|z) := \sum_{n \geq 1} \mathbb{P}\big[X_n = w_1, \forall m \in \{1, \ldots, n-1\} : X_m \neq w_1 \big| X_0 = w_1\big] \cdot z^n.$$

By conditioning on the last visit to w_1, an important relation between these functions is given by

$$G(w_1, w_2|z) = G(w_1, w_1|z) \cdot L(w_1, w_2|z).$$

Denote by R_w the radius of convergence of $G(w, w|z)$, $w \in \mathcal{L}$. If $R_w > 1$ then

$$G(w, w|1) \leq \frac{1}{1 - \frac{1}{R_w}}; \tag{3.1}$$

indeed, since $G(w, w|z) = \big(1 - U(w, w|z)\big)^{-1}$ it must be that $U(w, w|z) < 1$ for all $0 < z < R_w$; moreover, $U(w, w|0) = 0$, $U(w, w|z)$ is continuous, strictly increasing and strictly convex for $0 < z < R_w$, so we must have $U(w, w|1) \leq 1/R_w$ which yields (3.1).

In the following we introduce further generating functions, which also have been used in [6]. Define for $a, b, c, d, e \in \mathcal{A}$ and real $z > 0$

$$H(ab, c|z) := \sum_{n \geq 1} \mathbb{P}\big[X_n = c, \forall m < n : |X_m| > 1 \big| X_0 = ab\big] \cdot z^n$$

and

$$\bar{L}(ab, cde|z) \quad := \quad \sum_{n \geq 1} \mathbb{P}\big[X_n = cde, \forall m \in \{1, \ldots, n\} : |X_m| \geq 3 \big| X_0 = ab\big] \cdot z^n,$$

$$\bar{G}(ab, cd|z) \quad := \quad \sum_{n \geq 0} \mathbb{P}\big[X_n = cd, \forall m \in \{1, \ldots, n\} : |X_m| \geq 2 \big| X_0 = ab\big] \cdot z^n.$$

We write $\bar{L}(ab, cde) := \bar{L}(ab, cde|1)$. These generating functions can be computed in two steps: first, one solves the following system of equations:

$$\begin{aligned}
H(ab, c|z) \quad = \quad & p(ab, c) \cdot z + \sum_{de \in \mathcal{A}^2} p(ab, de) \cdot z \cdot H(de, c|z) \\
& + \sum_{def \in \mathcal{A}^3} p(ab, def) \cdot z \cdot \sum_{g \in \mathcal{A}} H(ef, g|z) \cdot H(dg, c|z);
\end{aligned} \tag{3.2}$$

compare with [12] and [6]. The system (3.2) consists of equations of quadratic order, and therefore the functions $H(\cdot, \cdot | z)$ are algebraic, if the transition probabilities are algebraic. We now get the functions $\bar{G}(ab, cd|z)$ by solving the following linear system of equations:

$$
\overline{G}(ab, cd|z) = \delta_{ab}(cd) + \sum_{c_1 d_1 \in \mathcal{A}^2} p(ab, c_1 d_1) \cdot z \cdot \overline{G}(c_1 d_1, cd|z) +
$$
$$
+ \sum_{c_1 d_1 e_1 \in \mathcal{A}^3} p(ab, c_1 d_1 e_1) \cdot z \cdot \sum_{f \in \mathcal{A}} H(d_1 e_1, f|z) \cdot \overline{G}(c_1 f, cd|z).
$$

Finally, we get

$$
\bar{L}(ab, cde|z) = \sum_{d_1 e_1 \in \mathcal{A}^2} p(ab, cd_1 e_1) \cdot z \cdot \bar{G}(d_1 e_1, ef|z).
$$

We remark that we implicitely took into account the assumption $\mathcal{L} = \mathcal{A}^*$; if $\mathcal{L} \subset \mathcal{A}^*$ one has to restrict these definitions and systems of equations to the terms which may occur. Moreover, one can compute the Green functions of the form $G(o, abc|z)$ by solving

$$
G(w_1, w_2|z) = \delta_{w_1}(w_2) + \sum_{w_3 \in \mathcal{L} : |w_3| \leq 3} p(w_1, w_3) \cdot z \cdot G(w_3, w_2|z) +
$$
$$
+ \mathbb{1}_3(w_1) \cdot \sum_{cde \in \mathcal{A}^3} p(w_1[2]w_1[3], cde) \cdot z \cdot \sum_{f \in \mathcal{A}} H(de, f|z) \cdot G(w_1[1]cf, w_2|z),
$$

where $w_1, w_2 \in \mathcal{L}$ with $|w_1|, |w_2| \leq 3$ and $\mathbb{1}_3(w_1) = 1$, if $|w_1| = 3$, and $\mathbb{1}_3(w_1) = 0$ otherwise.

We also define for $ab \in \mathcal{A}^2$:

$$
\xi(ab) := \sum_{cde \in \mathcal{A}^3} p(ab, cde) \cdot \Big(1 - \sum_{f \in \mathcal{A}} H(de, f|1)\Big).
$$

This is the probability of starting at a word $wab \in \mathcal{L}$, where $w \in \mathcal{A}^*$, such that the first step goes to a word of length $|wab| + 1$ with no further future visits of words of length $|wab|$ or smaller. We define a "length function" on $\mathcal{L}$ by

$$
l(x_1 \ldots x_n) := -\log L(o, x_1 \ldots x_n) \quad \text{for } x_1 x_2 \ldots x_n \in \mathcal{L}. \tag{3.3}
$$

For $n \geq 5$, the terms $L(o, x_1 \ldots x_n)$ can be rewritten as

$$
\sum_{b_1 c_1 \in \mathcal{A}^2} L(o, x_1 b_1 c_1 | 1) \sum_{\substack{b_2, \ldots, b_{n-3} \in \mathcal{A}, \\ c_2, \ldots, c_{n-3} \in \mathcal{A}}} \prod_{i=2}^{n-3} \bar{L}(b_{i-1} c_{i-1}, x_i b_i c_i) \cdot \bar{L}(b_{n-3} c_{n-3}, x_{n-2} x_{n-1} x_n); \tag{3.4}
$$

each path from o to $x_1 \ldots x_n$ is decomposed to the last times when the sets $\mathcal{A}^3, \mathcal{A}^4, \ldots, \mathcal{A}^{n-1}$ are visited.

4. Cones

In this section we introduce the structure of cones in our setting. A path in $\mathcal{L}$ is a sequence of words $[w_0, w_1, \ldots, w_m]$ in $\mathcal{L}$ such that $\mathbb{P}_{w_{i-1}}[X_1 = w_i] > 0$ for all $1 \leq i \leq m$. For $n \in \mathbb{N}$,

define $\mathcal{L}_{\geq n} := \{w \in \mathcal{L} \mid |w| \geq n\}$. For any $w \in \mathcal{L}$ with $|w| \geq 2$, we define the *cone* rooted at w as

$$C(w) := \{\bar{w} \in \mathcal{L}_{\geq |w|} \mid \exists \text{ path } [w, w_1, \dots, w_{m-1}, \bar{w}] \text{ with } m \in \mathbb{N}, w_1, \dots, w_{m-1} \in \mathcal{L}_{\geq |w|}\}.$$

By the above made weak symmetry assumption, for $w_1, w_2 \in \mathcal{L}$ with $|w_1| = |w_2|$, we have $C(w_1) = C(w_2)$ whenever there is a positive probability path from w_1 to w_2 in $\mathcal{L}_{\geq |w_1|}$.

By suffix-irreducibility, for all $cd \in \mathcal{A}^2$, each cone $C(wab)$, where $w \in \mathcal{A}^*$ and $ab \in \mathcal{A}^2$, has a subcone $C(wxcd) \subseteq C(w)$ with a suitable choice of $x \in \mathcal{A}^* \setminus \{o\}$. We say that two cones $C(w_1 \dots w_m)$ and $C(y_1 \dots y_n)$ are isomorphic if $C(w_{m-1}w_m) = C(y_{n-1}y_n)$, that is, two isomorphic cones differ only by different prefixes. In particular, there is a natural 1-to-1 correspondence of paths inside $C(w_1 \dots w_m)$ and paths in $C(y_1 \dots y_n)$ where obviously each pair of corresponding paths has the same probability. Since the transition probabilities depend only on the last two letters of the current word, there are only finitely many different cone types up to isomorphisms. We identify the different cone types by two-lettered words $ab \in \mathcal{A}^2$, and write $\tau(C(w)) = ab$ for its cone type, where ab are the last two letters of w. For each isomorphism class of cone types we fix some ab representing its cone type. Let $J \subseteq \mathcal{A}^2$ be the set of different cone types. The *boundary* $\partial C(w)$ of $C(w)$ is given by all words $w_0 \in C(w)$ with $|w_0| = |w|$. An important property is the following one: if $C(w_1)$ and $C(w_2)$ are two isomorphic cones with $w_0 ab \in \partial C(w_1)$, then there is $\bar{w}_0 \in \mathcal{A}^*$ such that $\bar{w}_0 ab \in \partial C(w_2)$.

Now we make the *non-singular covering assumption* that each cone $C(wa_0 b_0)$, $w \in \mathcal{A}^*$, $a_0 b_0 \in \mathcal{A}^2$, contains two proper disjoint subcones, that is, we assume that there are subcones of the form $C(ww_1 a_1 b_1), C(ww_2 a_2 b_2) \subsetneq C(wa_0 b_0)$ with $w_i \in \mathcal{A}^* \setminus \{o\}$, $a_i b_i \in \mathcal{A}^2$ and $C(ww_1 a_1 b_1) \cap C(ww_2 a_2 b_2) = \emptyset$. We refer to the remarks at the end of this section if this property does not hold. The next task is to cover (up to a finite complement) any cone $C(w)$ by a finite number of pairwise disjoint subcones $C_1, \dots, C_{r(w)}$ such that

$$\bigcup_{i=1}^{r(w)} \tau(C_i) = J \quad \text{and} \quad \left| C \setminus \bigcup_{i=1}^{r(w)} C_i \right| < \infty,$$

that is, among these subcones every cone type appears. We now show how to construct this covering. Suppose we are given a cone $C(wa_0 b_0)$ with $w \in \mathcal{A}^*$ and $a_0 b_0 \in \mathcal{A}^2$. Inside this cone we find subcones of the form $C(ww_0 ab)$ for each $ab \in \mathcal{A}^2$ with suitable $w_0 \in \mathcal{A}^* \setminus \{o\}$. Furthermore, we can choose these subcones in a way such that they are not contained in each other, that is, $C(ww_1 a_1 b_1) \not\subseteq C(ww_2 a_2 b_2)$ for all these chosen cones of all different types: indeed, since we assume existence of a non-singular covering of $C(w)$ by subcones one can walk from w inside $\mathcal{L}_{\geq |w|}$ to words $ww_1 a_1 b_1$ and $ww_2 a_2 b_2$, where $w_1, w_2 \in \mathcal{A}^*$, $a_i b_i \in \mathcal{A}^2$ and $C(ww_1 a_1 b_1) \cap C(ww_2 a_2 b_2) = \emptyset$. Then we have found a subcone of type $\tau(C(a_1 b_1))$, and we search for other cone types in the subcone $C(ww_2 a_2 b_2)$. Obviously, a subcone in $C(ww_2 a_2 b_2)$ does not intersect $C(ww_1 a_1 b_1)$. Iterating this step leads to subcones in $C(w)$ of all different types which do not intersect each other. After we have found non-intersecting subcones of all types in $C(w)$ we cover this cone by further subcones, which are not intersecting the above chosen subcones, such that the difference of $C(w)$ and the union of subcones is finite. This is, for instance, done by taking all cones rooted at words

$v \in C(w)$, where v is at the same distance (that is, minimal length of a path)to $\partial C(w)$ as the subcone of maximal distance to w and where v is not contained in any of the above chosen subcones yet. See Figure 1.
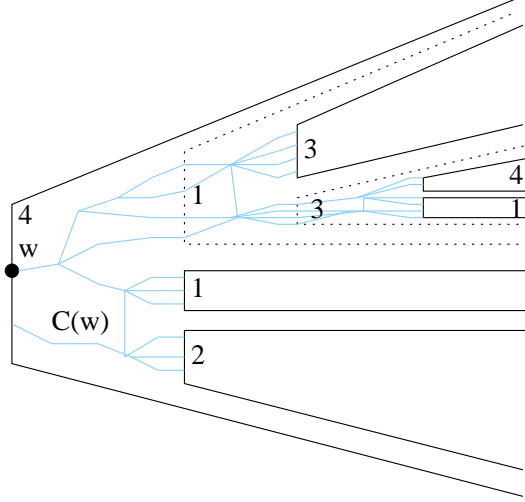

FIGURE 1. Covering of cones by subcones: the numbers represent the different cone types; the cones with entire boundary lines belong to the covering.

Let us remark that, for each cone type, we fix such a covering, such that the covering of $C(w)$ does not depend on the choice of the specific root w on the boundary of $C(w)$: fix a covering for $C(ab)$, $ab \in \mathcal{A}^2$; if $w = w_0 a_1 b_1 \in \mathcal{L}$ with $\tau(C(w)) = ab$ then we fix the covering of $C(w) = C(w_0 ab)$ which is inherits the covering from $C(ab)$. This is well-defined since the covering of a cone depends only on the relative location of its subcones in its interior.

We can also cover $\mathcal{L}$ (up to a finite set) by a finite number of non-self-containing subcones, where each cone type appears. To this end, we just apply the algorithm explained above and take cones of the form $C(w)$ with $|w| \geq 2$. We denote by $C_1^{(0)}, \ldots, C_{n_0}^{(0)}$ the covering of $\mathcal{L}$, which contains all types and whose complement is finite.

Now we explain how to proceed if every cone contains no two disjoint subcones. This case may, in particular, occur if $\mathcal{L}$ is a proper subset of $\mathcal{A}^*$. For $ab, cd \in \mathcal{A}^2$, observe that $cd \in C(ab)$ if and only if $ab \in C(cd)$. This implies that $C(w) = \{v \in \mathcal{L} \mid |v| \geq |w|\}$ and, in particular, that there is only just one cone type. We can then cover $C(w)$ by the subcone $C(w_1)$ for any $w_1 \in \mathcal{L}$ with $|w_1| = |w| + 1$ and $p(w, w_1) > 0$. One can show that in this case the random walk converges almost surely to a deterministic infinite word and that the support of the random walk is a proper subset of $\mathcal{A}^*$. In order to see this, assume that the random walk tends with some positive probabilities to some infinite words with prefixes $wabc$ and $wdef$, where $w \in \mathcal{A}^*$, $a, b, c, d, e, f \in \mathcal{A}$ with $a \neq d$. Since $C(wabc) \cap C(wdef) = \emptyset$ it must be that the random walk enters either $C(wabc)$ or $C(wdef)$ on its way to infinity due to the assumption of singular covering. That is, the letter a is deterministic, and by induction the infinite limiting word is deterministic

We call the random walk *expanding* if each cone contains two disjoint subcones. The results below depend not on the fact if the random walk is expanding or not. At the end, however, we will see that the non-expanding case leads to zero entropy.

5. EXIT TIMES

In this section we prove a law of large numbers, which turns out to be the asymptotic entropy in the later section. For this purpose, we define exit times (compare with [6]), for which we derive a law of large numbers. Throughout this section, we use the following notations: $w_0, w_1, w_2 \in \mathcal{A}^* \setminus \{o\}$ and $a, b, c, d, a_1, b_1, a_2, b_2, \cdots \in \mathcal{A}$.

5.1. **Exit Time Process.** We define the following *exit times*. Let $\mathbf{e}_0$ be the last time at which the random walk visits $\bigcup_{i=1}^{n_0} \partial C_i^{(0)}$ and stays in one of the cones $C_1^{(0)}, \ldots, C_{n_0}^{(0)}$ afterwards forever, that is,

$$\mathbf{e}_0 := \sup\{m \in \mathbb{N}_0 \mid X_m \in \partial C_i^{(0)}, \forall n > m : X_n \in C_i^{(0)} \setminus \partial C_i^{(0)} \text{ for any } i \in \{1, \ldots, n_0\}\}.$$

Inductively, if $X_{\mathbf{e}_k} = w$ and $C(w)$ has a covering (determined only by the type of $C(w)$) consisting of subcones $C_1^{(k)}, \ldots, C_{r(w)}^{(k)}$ as explained in Section 4, then

$$\mathbf{e}_{k+1} := \sup\{m > \mathbf{e}_k \mid X_m \in \partial C_i^{(k)}, \forall n > m : X_n \in C_i^{(k)} \setminus \partial C_i^{(k)} \text{ for any } i \in \{1, \ldots, r(w)\}\}.$$

Observe that X_n, $n \geq \mathbf{e}_k$, has the prefix w_0 if $X_{\mathbf{e}_k} = w_0 ab$. Define the *relative increment* between two exit times as follows: set $\mathbf{W}_0 := X_{\mathbf{e}_0}$; for $k \geq 1$: if $X_{\mathbf{e}_{k-1}} = w_0 ab$ and $X_{\mathbf{e}_k} = w_0 w_1 cd$, then $\mathbf{W}_k := w_1 cd$ for $k \geq 1$. Since we have only finitely many different cone types and the subcones of coverings of any cone C are nested at uniformly bounded distance (w.r.t. minimal path lengths) to ∂C, the random variables $\mathbf{W}_k$ can take only finitely many different values.

For $x \in \mathcal{L}$, define

$$\mathcal{S}(x) := \bigcup_{i=1}^{r(x)} \partial C_i,$$

where $C_1, \ldots, C_{r(x)}$ is a covering of $C(x)$. Furthermore, define for $x = x_1 \ldots x_n \in \mathcal{L}$ and $y = x_1 \ldots x_{n-2} x'_{n-1} x'_n \ldots x'_{n+d} \in C(x)$ with $d = d(x, y) := |y| - |x|$

$$\mathbb{L}(x, y) := \sum_{n \geq 0} \mathbb{P}\left[X_n = y, \forall m \in \{1, \ldots, n\} : X_m \in C(x) \setminus \partial C(x) \mid X_0 = x\right]$$

$$= \sum_{\substack{y_1, \ldots, y_{d-1} : y_i \in \mathcal{A}^3 : \\ y_i[1] = x'_{n-2+i}}} \bar{L}(x_{n-1} x_n, y_1) \cdot \bar{L}(y_1[2] y_1[3], y_2) \cdot \ldots \cdot \bar{L}(y_{d-1}, x'_{n+d-2} x'_{n+d-1} x'_{n+d}).$$

Obviously, $\mathbb{L}(x, y)$ depends on x only by its last two letters.

Proposition 5.1. *The process $(\mathbf{W}_k)_{k \geq 1}$ is a positiv recurrent Markov chain with transition probabilities*

$$q(x, y) := \begin{cases} \frac{\xi([y])}{\xi([x])} \mathbb{L}(x, y), & \text{if } y \in \mathcal{S}(x), \\ 0, & \text{otherwise.} \end{cases}$$

Proof. Let be $w_0, \ldots, w_{k+1} \in \mathcal{A}^* \setminus \{o\}$ such that $w_{i+1} \in S(w_i)$ for all $i \in \{0, \ldots, k\}$ and $w_0 \in \bigcup_{j=1}^{n_0} \partial C_j^{(0)}$. We set $x_0 := w_0$ and inductively: if $x_{k-1} = y_{k-1}a_{k-1}b_{k-1}$ with $y_{k-1} \in \mathcal{A}^*$ and $a_{k-1}b_{k-1} \in \mathcal{A}^2$ then set $x_k := y_{k-1}w_k$. Then:

$$\mathbb{P}\big[\mathbf{W}_0 = w_0, \ldots, \mathbf{W}_k = w_k\big] = \mathbb{P}\big[X_{\mathbf{e}_0} = x_0, \ldots, X_{\mathbf{e}_k} = x_k\big]$$
$$= \ G(o, w_0|1) \cdot \mathbb{L}(w_0, w_1)\mathbb{L}(w_1, w_2) \cdot \ldots \cdot \mathbb{L}(w_{k-1}, w_k) \cdot \xi([w_k]).$$

Consider

$$\mathbb{P}\big[\mathbf{W}_{k+1} = w_{k+1} \mid \mathbf{W}_0 = w_0, \ldots, \mathbf{W}_k = w_k\big]$$
$$= \ \frac{\mathbb{P}\big[\mathbf{W}_0 = w_0, \ldots, \mathbf{W}_k = w_k, \mathbf{W}_{k+1} = w_{k+1}\big]}{\mathbb{P}\big[\mathbf{W}_0 = w_0, \ldots, \mathbf{W}_k = w_k\big]}$$
$$= \ \frac{G(o, w_0|1) \cdot \mathbb{L}(w_0, w_1)\mathbb{L}(w_1, w_2) \cdot \ldots \cdot \mathbb{L}(w_{k-1}, w_k) \cdot \mathbb{L}(w_k, w_{k+1}) \cdot \xi([w_{k+1}])}{G(o, w_0|1) \cdot \mathbb{L}(w_0, w_1)\mathbb{L}(w_1, w_2) \cdot \ldots \cdot \mathbb{L}(w_{k-1}, w_k) \cdot \xi([w_k])} = q(x, y).$$

Since there are only finitely many different values for $\mathbf{W}_k$ positive recurrence follows due to suffix-irreducibility, which implies irreducibility of the process $\big(\mathbf{W}_k\big)_{k \geq 1}$. $\qquad\square$

The random variables $\mathbf{W}_k$, $k \geq 1$, can take values in

$$\mathcal{W}_0 := \big\{w \in \mathcal{A}^* \big| \mathbb{P}[\mathbf{W}_2 = w \mid \mathbf{W}_1 = w_0ab] > 0 \text{ for any } w_0 \in \mathcal{A}^*, ab \in \mathcal{A}^2\big\}.$$

Observe that the transition probabilities depend on x only by its last two letters.

Lemma 5.2. *We have* $\mathrm{supp}(\mathbb{P}[\mathbf{W}_1 = \cdot]) = \mathcal{W}_0$.

Proof. Let $y = a_1b_1w_ya_2b_2 \in \mathcal{W}_0$ with $w_y \in \mathcal{A}^*$ (we omit the special case $y = a_1a_2b_2$ which follows analogously). Then there is $\bar{a}_1\bar{b}_1 \in \mathcal{A}^2$ with $\mathbb{L}(\bar{a}_1\bar{b}_1, y) > 0$ and $\xi(a_2b_2) > 0$. By construction of our coverings there is some $w_0 \in \mathcal{A}^*$ with $w_0\bar{a}_1\bar{b}_1 \in \bigcup_{i=1}^{n_0} \partial C_i^{(0)}$. Choose $m \in \mathbb{N}$ such that $p^{(m)}(o, w_0\bar{a}_1\bar{b}_1) > 0$. Then:

$$\mathbb{P}[\mathbf{W}_1 = y] \geq p^{(m)}(o, w_0\bar{a}_1\bar{b}_1) \cdot \mathbb{L}(\bar{a}_1\bar{b}_1, y) \cdot \xi(a_2b_2) > 0.$$

$\qquad\square$

For sake of better identification of the cones, we now switch to a more suitable representation of cones and coverings. We identify the different cone types by numbers $\mathcal{I} := \{1, \ldots, r\} \subset \mathbb{N}$. If $C(w)$ is a cone of type i, then the covering of $C(w)$ has n_j subcones of type j. We denote these subcones by $C_{j_{i,1}} = C_{j_{i,1}}(w), \ldots, C_{j_{i,n_j}} = C_{j_{i,n_j}}(w)$ or identify them just by $j_{i,1}, \ldots, j_{i,n_j}$, which correspond to the cones of type j with different locations inside $C(w)$. We will sometimes omit the root w in the notation of the subcones when it will be clear from the context and only the relative positon of a subcone in some given cone will be important. If $\tau(C(X_{\mathbf{e}_{k-1}})) = i$ and $X_{\mathbf{e}_k} \in \partial C_{j_{i,l}}(X_{\mathbf{e}_{k-1}})$, then we set $\mathbf{i}_k := j_{i,l}$.

At this point we recall the relation between $\mathbf{W}_k$ and $X_{\mathbf{e}_k}$: if $X_{\mathbf{e}_0} = w_0a_0b_0$ and $\mathbf{W}_1 = w_1a_1b_1$ then $X_{\mathbf{e}_1} = w_0w_1a_1b_1$; in general, if $X_{\mathbf{e}_{k-1}} = wa_{k-1}b_{k-1}$ and $\mathbf{W}_k = w_ka_kb_k$ then $X_{\mathbf{e}_k} = ww_ka_kb_k$. That is, there is a natural bijection of trajectories of $(\mathbf{W}_k)_{k \in \mathbb{N}}$ and $(X_{\mathbf{e}_k})_{k \in \mathbb{N}}$. In particular, the value of $\mathbf{W}_k$ determines the value of $\mathbf{i}_k$ uniquely. For a better visualization of the values $\mathbf{i}_k := j_{i,l}$, see Figure 2. In other words, the random variables
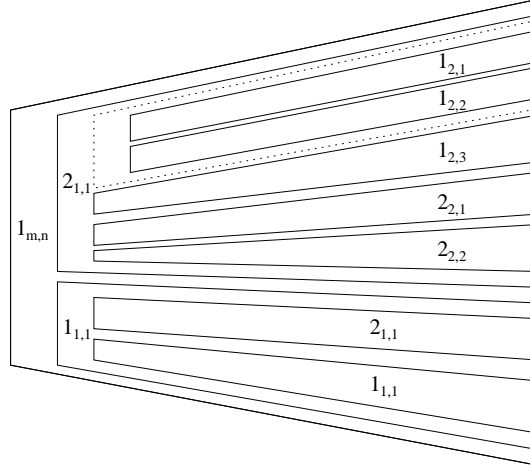
FIGURE 2. Numbering of subcones

$\mathbf{i}_k$ collect the different cones which are entered successively by the random walk $(X_n)_{n\in\mathbb{N}_0}$ on its way to infinity, while the $\mathbf{W}_k$'s keep, in addition, the information where the single subcones are finally entered.

For $s,t\in\mathcal{I}$, let $n(s,t)$ denote the number of different cones of type t in the covering of a cone of type s. Define

$$\mathcal{W} := \left\{(j_{m,n},x)\,\middle|\, \begin{array}{l} j,m\in\mathcal{I}, \tau(C(x[1]x[2])=m, \tau(C(x))=j, \\ 1\le n\le n(m,j), x\in\partial C_{j_{m,n}}(x[1]x[2])\cap\mathcal{W}_0 \end{array}\right\}.$$

In other words, $(j_{m,n},x)\in\mathcal{W}$ if $x=a_0b_0w_0ab\in\mathcal{W}_0$ with $\tau(C(a_0b_0))=m$ and $C(x)$ is the n-th cone of type j inside $C(a_0b_0)$. Furthermore, define

$$\mathcal{W}_\pi := \big\{(s,t_n)\big| s,t\in\mathcal{I}, 1\le n\le n(s,t)\big\}.$$

That is, t_n corresponds to the n-th cone of type t in a covering of a cone of type s.

The process $\big((\mathbf{i}_k,\mathbf{W}_k)\big)_{k\in\mathbb{N}}$ with state space $\mathcal{W}$ is also a positive recurrent Markov chain since the values of $\mathbf{i}_k$ are uniquely determined by the values of $\mathbf{W}_k$ and the process $(\mathbf{W}_k)_{k\in\mathbb{N}}$ is a Markov chain. Moreover, for $(i_{k,l},w_{k-1}),(j_{m,n},w_k)\in\mathcal{W}$, the transition probabilities are given by

$$\mathbb{P}\Big[(\mathbf{i}_k,\mathbf{W}_k)=(j_{m,n},w_k)\Big|(\mathbf{i}_{k-1},\mathbf{W}_{k-1})=(i_{k,l},w_{k-1})\Big] = \begin{cases} q(w_{k-1},w_k), & \text{if } m=i, \\ 0, & \text{if } m\neq i. \end{cases}$$

In particular, the transition matrix of $\big((\mathbf{i}_k,\mathbf{W}_k)\big)_{k\in\mathbb{N}}$ has zero entries. In order to apply the result of [10, Theorem 1.1] for getting the analytic behaviour of the entropy later we have to adapt the Markov chain in order to obtain a transition matrix without zeroes.

The process $(\mathbf{i}_k)_{k\in\mathbb{N}}$ is, in general, not a Markov chain because it can be seen as a projection of the process $(\mathbf{W}_k)_{k\in\mathbb{N}}$.

Define the following projection for $(i_{k,l}, w_1), (j_{m,n}, w_2) \in \mathcal{W}$:

$$\pi\big((i_{k,l}, w_1), (j_{m,n}, w_2)\big) := \begin{cases} (i, j_{i,n}) =: (i, j_n), & \text{if } m = i, \\ (i, j_{i,1}) = (i, j_1), & \text{if } m \neq i. \end{cases}$$

Here, j_l represents the l-th cone of type j in a cone of type i, namely the cone represented by $j_{i,l}$. We now define the *hidden Markov chain* $(\mathbf{Y}_k)_{k \in \mathbb{N}}$ by

$$\mathbf{Y}_k := \pi\big((\mathbf{i}_k, \mathbf{W}_k), (\mathbf{i}_{k+1}, \mathbf{W}_{k+1})\big).$$

In other words, $(\mathbf{Y}_k)_{k \in \mathbb{N}}$ traces the way to infinity in terms of which subcones are entered successively *without* distinguishing which of the hit boundary points are the exit time points $X_{\mathbf{e}_k}$.

Morever, observe that $\text{supp}(\mathbb{P}[\mathbf{Y}_1 = \cdot]) = \mathcal{W}_\pi$ since – by construction of the coverings – for any $(s, t_n) \in \mathcal{W}_\pi$ there is $w_0 ab \in \bigcup_{i=1}^{n_0} \partial C_i^{(0)}$ and $x \in \mathcal{W}_0$ with $\tau(C(x)) = s$ and $\mathbb{P}[X_{\mathbf{e}_1} = w_0 x \mid X_{\mathbf{e}_0} = w_0 ab] > 0$ and there is $y \in \partial C_{t_{s,n}}([x])$ with $q(x, y) > 0$ such that

$$\begin{aligned} \mathbb{P}[\mathbf{Y}_1 = (s, t_n)] &\geq \mathbb{P}\big[X_{\mathbf{e}_0} = w_0 ab, X_{\mathbf{e}_1} = w_0 x, \mathbf{W}_2 = y\big] \\ &\geq \mathbb{P}[X_{\mathbf{e}_0} = w_0 ab] \cdot \mathbb{P}[X_{\mathbf{e}_1} = w_0 x \mid X_{\mathbf{e}_0} = w_0 ab] \cdot q(x, y) > 0. \end{aligned}$$

5.2. **Modified Exit Time Process.** The aim of this subsection is the construction of a Markov chain related to the exit time process $(\mathbf{i}_k, \mathbf{W}_k)_{k \in \mathbb{N}}$ such that the transition matrix has strictly positive entries and the modified process leads under π to the same hidden Markov chain for almost every trajectory.

Consider the two subcones $C_{j_{i,1}} \subset C(a_1 b_1)$ and $C_{j_{k,l}} \subset C(a_2 b_2)$ belonging to coverings of the bigger cones with $\tau(C(a_1 b_1)) = i$ and $\tau(C(a_2 b_2)) = k$. Assume that $y_0 ab \in \partial C_{j_{k,n}}$. Since both cones are isomorphic, there is unique $\bar{y}_0 = \bar{y}_0^{[i,j,ab]} \in \mathcal{A}^*$ such that $\bar{y}_0 ab \in \partial C_{j_{i,1}}$; see Figure 3. In the following we will use this notation $\bar{y}_0 = \bar{y}_0^{[i,j,ab]}$ for describing this
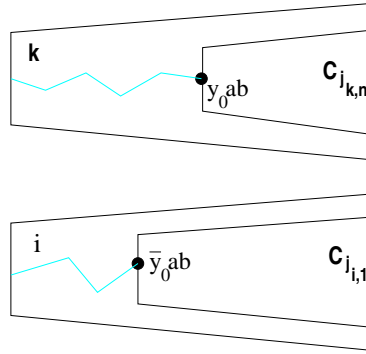

FIGURE 3. Prefix replacement

replacement.

For $i, j \in \mathcal{I}$ and $ab \in \mathcal{A}^2$ with $\tau(C(ab)) = j$ we write

$$\#\{j_{s,t} \mid s \neq i\} = \big|\{(j_{s,t}, xab) \in \mathcal{W} \mid s \in \mathcal{I} \setminus \{i\}, 1 \leq t \leq n(s,j)\}\big|,$$

which is independent from the specific choice of ab. Let $(i_{k,l}, x), (j_{m,n}, y_0 ab) \in \mathcal{W}$. Define the following transition probabilities on $\mathcal{W}$:

$$\hat{q}\big((i_{k,l}, x), (j_{m,n}, y_0 ab)\big) := \begin{cases} \frac{1}{\#\{j_{s,t} | s \neq i\} + 1} \frac{\xi(ab)}{\xi([x])} \mathbb{L}(x, y_0 ab), & \text{if } m = i \wedge n = 1, \\ \frac{\xi(ab)}{\xi([x])} \mathbb{L}(x, y_0 ab), & \text{if } m = i \wedge n \geq 2, \\ \frac{1}{\#\{j_{s,t} | s \neq i\} + 1} \frac{\xi(ab)}{\xi([x])} \mathbb{L}(x, \bar{y}_0^{[i,j,ab]} ab), & \text{if } m \neq i. \end{cases}$$

Observe that the transitions depend on x only by its last two letters. It is easy to see that these transition probabilities define a Markov chain (inherited from the process $(\mathbf{i}_k, \mathbf{W}_k)_{k \in \mathbb{N}}$): each step from $(i_{k,l}, x)$ to $(j_{m,n}, y_0 ab)$ either behaves according to $q(\cdot, \cdot)$ (case $m = i$ and $n \geq 2$) or steps from $(i_{k,l}, x)$ to $(j_{i,1}, y_0 ab)$ (when seen as a step of the process $(\mathbf{i}_k, \mathbf{W}_k)_{k \in \mathbb{N}}$) are split up into different equally likely paths $(i_{k,l}, x)$ to $(j_{m,n}, y_0 ab)$ with $m \neq i$ or $m = i \wedge n = 1$; since $\hat{q}(\cdot, \cdot)$ depends on its first argument only by i (and not by k and l), it follows from $q(\cdot, \cdot)$ that $\hat{q}(\cdot, \cdot)$ describes also a random walk. Moreover, the corresponding transition matrix has strictly positive entries. By suffix-irreducibility and Proposition 5.1, the matrix $\widehat{Q} = \big(\hat{q}((i_{k,l}, x), (j_{m,n}, y))\big)$ is stochastic, and governs a positiv recurrent Markov chain $(\hat{\mathbf{i}}_k, \mathbf{x}_k)_{k \in \mathbb{N}}$ with invariant probability measure ν. The initial distribution of $(\hat{\mathbf{i}}_1, \mathbf{x}_1)$ is given by $\hat{\mu}_1$, defined as

$$\hat{\mu}_1(i_{m,n}, x) := \mathbb{P}[(\mathbf{i}_1, \mathbf{W}_1) = (i_{m,n}, x)] = \mathbb{P}[\tau(C(X_{\mathbf{e}_0})) = m, \mathbf{W}_1 = x] > 0$$

for $(i_{m,n}, x) \in \mathcal{W}$. If we equip the process with the invariant probability measure ν as initial distribution we write $\big(\hat{\mathbf{i}}_k^{(\nu)}, \mathbf{x}_k^{(\nu)}\big)_{k \in \mathbb{N}}$.

Then the process $\big((\hat{\mathbf{i}}_k, \mathbf{x}_k), (\hat{\mathbf{i}}_{k+1}, \mathbf{x}_{k+1})\big)_{k \in \mathbb{N}}$ is again a positiv recurrent Markov chain with transition matrix $\widehat{Q}_2$ (arising from $\widehat{Q}$) and invariant probability measure denoted by ν_2. Once again, if we equip this process with the initial distribution ν_2, which arises in a natural way from ν, then we write $\big((\hat{\mathbf{i}}_k^{(\nu)}, \mathbf{x}_k^{(\nu)}), (\hat{\mathbf{i}}_{k+1}^{(\nu)}, \mathbf{x}_{k+1}^{(\nu)})\big)_{k \geq 1}$.

We now define two hidden Markov chains $(\mathbf{Z}_k^{(\nu)})_{k \in \mathbb{N}}$ and $(\widehat{\mathbf{Y}}_k)_{k \in \mathbb{N}}$ by

$$\mathbf{Z}_k^{(\nu)} := \pi\big((\hat{\mathbf{i}}_k^{(\nu)}, \mathbf{x}_k^{(\nu)}), (\hat{\mathbf{i}}_{k+1}^{(\nu)}, \mathbf{x}_{k+1}^{(\nu)})\big), \ \widehat{\mathbf{Y}}_k := \pi\big((\hat{\mathbf{i}}_k, \mathbf{x}_k), (\hat{\mathbf{i}}_{k+1}, \mathbf{x}_{k+1})\big).$$

That is, $(\mathbf{Z}_k^{(\nu)})_{k \in \mathbb{N}}$ and $(\widehat{\mathbf{Y}}_k)_{k \in \mathbb{N}}$ differ in their evolution only in their inital distributions. The crucial point now is the following proposition:

Proposition 5.3. *For all* $(s^{(1)}, t^{(1)}), \ldots, (s^{(n)}, t^{(n)}) \in \mathcal{W}_\pi$,

$$\mathbb{P}\big[\mathbf{Y}_1 = (s^{(1)}, t^{(1)}), \ldots, \mathbf{Y}_n = (s^{(n)}, t^{(n)})\big] = \mathbb{P}\big[\widehat{\mathbf{Y}}_1 = (s^{(1)}, t^{(1)}), \ldots, \widehat{\mathbf{Y}}_n = (s^{(n)}, t^{(n)})\big].$$

Proof. We prove the claim by induction on n. First, let $j, s \in \mathcal{I}$ and $t^{(1)} = j_m$ with $2 \leq m \leq n(s, j)$, and let $a_0 b_0, ab \in \mathcal{A}^2$ with $\tau(C(a_0 b_0)) = s$ and $\tau(C(ab)) = j$. If $C_{j,m}$ is the m-th cone of type j in the covering of $C(a_0 b_0)$ then there is unique $x_0 = x_0^{[s,j,m,ab]} \in \mathcal{A}^*$

with $x_0 ab \in \partial C_{j,m}$. With this notation we get:

$$
\begin{aligned}
\mathbb{P}\big[\mathbf{Y}_1 = (s, j_m), [\mathbf{W}_2] = ab\big] &= \sum_{(u_{k,l}, x) \in \mathcal{W}: u = s} \mathbb{P}[(\mathbf{i}_1, \mathbf{W}_1) = (s_{k,l}, x)] \cdot q(x, x_0 ab) \\
&= \sum_{(u_{k,l}, x) \in \mathcal{W}: u = s} \hat{\mu}_1(s_{k,l}, x) \hat{q}\big((s_{k,l}, x), (j_{s,m}, x_0 ab)\big) \\
&= \mathbb{P}\big[\widehat{\mathbf{Y}}_1 = (s, j_m), [\mathbf{x}_2] = ab\big].
\end{aligned}
$$

Now we turn to the case $t^{(1)} = j_1$. Once again, if $C_{j,1}$ is the first cone of type j in the covering of $C(a_0 b_0)$ then there is unique $x_0 = x_0^{[s,j,1,ab]} \in \mathcal{A}^*$ with $x_0 ab \in \partial C_{j,1}$. We get:

$$
\begin{aligned}
&\mathbb{P}\big[\widehat{\mathbf{Y}}_1 = (s, j_1), [\mathbf{x}_2] = ab\big] \\
&= \sum_{(u_{k,l}, x) \in \mathcal{W}: u = s} \hat{\mu}_1(s_{k,l}, x) \Big[\hat{q}\big((s_{k,l}, x), (t_{s,1}, x_0 ab)\big) + \sum_{\substack{(t_{p,q}, ya_1 b_1) \in \mathcal{W}: \\ p \neq s, a_1 b_1 = ab}} \hat{q}\big((s_{k,l}, x), (t_{p,q}, yab)\big) \Big] \\
&= \sum_{(u_{k,l}, x) \in \mathcal{W}: u = s} \hat{\mu}_1(s_{k,l}, x) \Big[\frac{q\big((s_{k,l}, x), (t_{s,1}, x_0 ab)\big)}{\#\{t_{k,l} \mid k \neq s\} + 1} + \sum_{\substack{(t_{p,q}, ya_1 b_1) \in \mathcal{W}: \\ p \neq s, a_1 b_1 = ab}} \frac{q\big((s_{k,l}, x), (t_{s,1}, x_0 ab)\big)}{\#\{t_{k,l} \mid k \neq s\} + 1} \Big] \\
&= \sum_{(u_{k,l}, x) \in \mathcal{W}: u = s} \mathbb{P}\big[(\mathbf{i}_1, \mathbf{W}_1) = (s_{k,l}, x)\big] \cdot q(x, x_0 ab) = \mathbb{P}\big[\mathbf{Y}_1 = (s, j_1), [\mathbf{W}_2] = ab\big].
\end{aligned}
$$

Now,

$$
\begin{aligned}
\mathbb{P}\big[\widehat{\mathbf{Y}}_1 = (s, t^{(1)})\big] &= \sum_{ab \in \mathcal{A}^2} \mathbb{P}\big[\widehat{\mathbf{Y}}_1 = (s, t^{(1)}), [\mathbf{x}_2] = ab\big] \\
&= \sum_{ab \in \mathcal{A}^2} \mathbb{P}\big[\mathbf{Y}_1 = (s^{(1)}, t^{(1)}), [\mathbf{W}_2] = ab\big] = \mathbb{P}\big[\mathbf{Y}_1 = (s^{(1)}, t^{(1)})\big].
\end{aligned}
$$

We now perform the induction step where we will use the equations from the initial step as induction assumptions. First, consider the case $t^{(n+1)} = j_m$ with $m \geq 2$; then for all $a_0 b_0, ab \in \mathcal{A}^2$ with $\tau(C(a_0 b_0)) = s^{(n+1)} =: s$ and $\tau(C(ab)) = j$ there is unique $x_0 = x_0^{[s,j,m,ab]} \in \mathcal{A}^*$ with $x_0 ab \in \partial C_{j_{s,m}}(a_0 b_0)$. Since we have an underlying Markov chain we obtain:

$$
\begin{aligned}
&\mathbb{P}\big[\widehat{\mathbf{Y}}_1 = (s^{(1)}, t^{(1)}), \dots, \widehat{\mathbf{Y}}_{n+1} = (s^{(n+1)}, t^{(n+1)}), [\mathbf{x}_{n+1}] = a_0 b_0, [\mathbf{x}_{n+2}] = ab\big] \\
&= \sum_{\substack{(u, v_k) \in \mathcal{W}_\pi: w_0 \in \mathcal{A}^* \\ v = s}} \mathbb{P}\big[\widehat{\mathbf{Y}}_1 = (s^{(1)}, t^{(1)}), \dots, \widehat{\mathbf{Y}}_n = (u, s_k), [\mathbf{x}_{n+1}] = w_0 a_0 b_0\big] \\
&\qquad \cdot \hat{q}\big((s_k, w_0 a_0 b_0), (j_{s,m}, x_0 ab)\big) \\
&= \mathbb{P}\big[\widehat{\mathbf{Y}}_1 = (s^{(1)}, t^{(1)}), \dots, \widehat{\mathbf{Y}}_n = (s^{(n)}, t^{(n)}), [\mathbf{x}_{n+1}] = a_0 b_0\big] \frac{\xi(ab)}{\xi(a_0 b_0)} \mathbb{L}(a_0 b_0, x_0 ab) \\
&= \mathbb{P}\big[\mathbf{Y}_1 = (s^{(1)}, t^{(1)}), \dots, \mathbf{Y}_n = (s^{(n)}, t^{(n)}), [\mathbf{W}_{n+1}] = a_0 b_0\big] \frac{\xi(ab)}{\xi(a_0 b_0)} \mathbb{L}(a_0 b_0, x_0 ab) \\
&= \mathbb{P}\big[\mathbf{Y}_1 = (s^{(1)}, t^{(1)}), \dots, \mathbf{Y}_{n+1} = (s^{(n+1)}, t^{(n+1)}), [\mathbf{W}_{n+1}] = a_0 b_0, [\mathbf{W}_{n+2}] = ab\big].
\end{aligned}
$$

Now we turn to the case $t^{(n+1)} = j_1$. Once again, if $C_{j,1}$ is the first cone of type j in the covering of $C(a_0 b_0)$ (of type s) then there is unique $x_0 = x_0^{[s,j,1,ab]} \in \mathcal{A}^*$ with $x_0 ab \in \partial C_{j,1}$. We get by distinguishing whether $t^{(n+1)} = j_1$ arises from $\mathbf{i}_{n+2} = j_{s,1}$ or $\mathbf{i}_{n+2} = j_{k,l}$ with $k \neq s$:

$$\mathbb{P}\big[\widehat{\mathbf{Y}}_1 = (s^{(1)}, t^{(1)}), \dots, \widehat{\mathbf{Y}}_{n+1} = (s^{(n+1)}, j_1), [\mathbf{x}_{n+1}] = a_0 b_0, [\mathbf{x}_{n+2}] = ab\big]$$

$$= \sum_{(u_{p,q}, w) \in \mathcal{W}: [w] = a_0 b_0} \mathbb{P}\big[\widehat{\mathbf{Y}}_1 = (s^{(1)}, t^{(1)}), \dots, \widehat{\mathbf{Y}}_n = (s^{(n)}, t^{(n)}), \hat{\mathbf{i}}_{n+1} = u_{p,q}, \mathbf{x}_{n+1} = w\big]$$

$$\cdot \Big(\hat{q}\big((u_{p,q}, w), (j_{s,1}, x_0 ab)\big) + \sum_{\substack{(t_{k,l}, y) \in \mathcal{W}: \\ t=j, k \neq s, [y]=ab}} \hat{q}\big((u_{p,q}, w), (j_{k,l}, y)\big)\Big)$$

$$= \mathbb{P}\big[\widehat{\mathbf{Y}}_1 = (s^{(1)}, t^{(1)}), \dots, \widehat{\mathbf{Y}}_n = (s^{(n)}, t^{(n)}), [\mathbf{x}_{n+1}] = a_0 b_0\big]$$

$$\cdot \Big[\frac{\xi(ab)}{\xi(a_0 b_0)} \frac{\mathbb{L}(a_0 b_0, x_0 ab)}{\#\{j_{k,l} \mid k \neq s\} + 1} + \sum_{\substack{(t_{p,q}, y) \in \mathcal{W}: \\ t=j, p \neq s, [y]=ab}} \frac{\xi(ab)}{\xi(a_0 b_0)} \frac{\mathbb{L}(a_0 b_0, x_0 ab)}{\#\{j_{k,l} \mid k \neq s\} + 1}\Big]$$

$$= \mathbb{P}\big[\widehat{\mathbf{Y}}_1 = (s^{(1)}, t^{(1)}), \dots, \widehat{\mathbf{Y}}_n = (s^{(n)}, t^{(n)}), [\mathbf{x}_{n+1}] = a_0 b_0\big] \frac{\xi(ab)}{\xi(a_0 b_0)} \mathbb{L}(a_0 b_0, x_0 ab)$$

$$= \mathbb{P}\big[\mathbf{Y}_1 = (s^{(1)}, t^{(1)}), \dots, \mathbf{Y}_n = (s^{(n)}, t^{(n)}), [\mathbf{W}_{n+1}] = a_0 b_0\big] \frac{\xi(ab)}{\xi(a_0 b_0)} \mathbb{L}(a_0 b_0, x_0 ab)$$

$$= \mathbb{P}\big[\mathbf{Y}_1 = (s^{(1)}, t^{(1)}), \dots, \mathbf{Y}_{n+1} = (s^{(n+1)}, j_1), [\mathbf{W}_{n+1}] = a_0 b_0, [\mathbf{W}_{n+2}] = ab\big].$$

Finally, we obtain:

$$\mathbb{P}\big[\widehat{\mathbf{Y}}_1 = (s^{(1)}, t^{(1)}), \dots, \widehat{\mathbf{Y}}_{n+1} = (s^{(n+1)}, t^{(n+1)})\big]$$

$$= \sum_{a_0 b_0, ab \in \mathcal{A}^2} \mathbb{P}\big[\widehat{\mathbf{Y}}_1 = (s^{(1)}, t^{(1)}), \dots, \widehat{\mathbf{Y}}_{n+1} = (s^{(n+1)}, t^{(n+1)}), [\mathbf{x}_{n+1}] = a_0 b_0, [\mathbf{x}_{n+2}] = ab\big]$$

$$= \sum_{a_0 b_0, ab \in \mathcal{A}^2} \mathbb{P}\big[\mathbf{Y}_1 = (s^{(1)}, t^{(1)}), \dots, \mathbf{Y}_{n+1} = (s^{(n+1)}, t^{(n+1)}), [\mathbf{W}_{n+1}] = a_0 b_0, [\mathbf{W}_{n+2}] = ab\big]$$

$$= \mathbb{P}\big[\mathbf{Y}_1 = (s^{(1)}, t^{(1)}), \dots, \mathbf{Y}_{n+1} = (s^{(n+1)}, t^{(n+1)})\big].$$

This finishes the proof. $\qquad\square$

The statement of the lemma can be said in other words: the process governed by $\widehat{Q}$ can be seen as a exit time process, where one has more subcones to enter (namely, the subcones of indices $j_{k,l}, k \neq i$, when being currently in a cone of type i), but under the projection π folds the process down to the same hidden Markov chain $(\mathbf{Y}_k)_{k \in \mathbb{N}}$, and it does not distinguish if $\hat{\mathbf{i}}_k = j_{i,1}$ or $\hat{\mathbf{i}}_k = j_{m,n}, m \neq i$.

Hence, the Markov chains $\big((\mathbf{i}_k, \mathbf{W}_k), (\mathbf{i}_{k+1}, \mathbf{W}_{k+1})\big)_{k \in \mathbb{N}}$ and $\big((\hat{\mathbf{i}}_k, \mathbf{x}_k), (\hat{\mathbf{i}}_{k+1}, \mathbf{x}_{k+1})\big)_{k \in \mathbb{N}}$ lead to the same hidden Markov chain in terms of probability. The important difference is that the transition matrix $\widehat{Q}$ has strictly positive entries, while this must not hold for the transition matrix of the chain $\big((\mathbf{i}_k, \mathbf{W}_k), (\mathbf{i}_{k+1}, \mathbf{W}_{k+1})\big)_{k \in \mathbb{N}}$.

5.3. **Entropy of the Hidden Markov Chain related to the Exit Time Process.** In this subsection we derive existence of the asymptotic entropy of the hidden Markov chains $(\widehat{\mathbf{Y}}_k)_{k\in\mathbb{N}}$ and $(\mathbf{Y}_k)_{k\in\mathbb{N}}$.

First, consider the hidden markov chain $(\mathbf{Z}_k)_{k\in\mathbb{N}}$: this process is stationary and ergodic since the underlying Markov chain $\left(\mathbf{i}_k^{(\nu)}, \mathbf{x}_k^{(\nu)}\right)_{k\in\mathbb{N}}$ is stationary and ergodic. Hence, there is a constant $H(\mathbf{Z}) \geq 0$ such that

$$\lim_{k\to\infty} -\frac{1}{k}\log\mathbb{P}[\mathbf{Z}_1 = \underline{s}_1, \dots, \mathbf{Z}_k = \underline{s}_k] = H(\mathbf{Z})$$

for almost every realisation $(\underline{s}_1, \underline{s}_2, \dots) \in \mathcal{W}_\pi^{\mathbb{N}}$ of $(\mathbf{Z}_k)_{k\in\mathbb{N}}$; see e.g. Cover and Thomas [4, Theorem 16.8.1]. We now deduce the same property for the process $(\widehat{\mathbf{Y}}_k)_{k\in\mathbb{N}}$.

Proposition 5.4. *For almost every realisation* $(\underline{s}_1, \underline{s}_2, \dots) \in \mathcal{W}_\pi^{\mathbb{N}}$ *of* $(\widehat{\mathbf{Y}}_k)_{k\in\mathbb{N}}$,

$$\lim_{k\to\infty} -\frac{1}{k}\log\mathbb{P}\big[\widehat{\mathbf{Y}}_1 = \underline{s}_1, \dots, \widehat{\mathbf{Y}}_k = \underline{s}_k\big] = H(\mathbf{Z}).$$

Proof. The processes $(\mathbf{Z}_k)_{k\in\mathbb{N}}$ and $(\widehat{\mathbf{Y}}_k)_{k\in\mathbb{N}}$ differ only in the inital distribution. Moreover, there are constants $c, C > 0$ such that

$$c \cdot \hat{\mu}_1(i_{m,n}, x) \leq \nu(i_{m,n}, x) \leq C \cdot \hat{\mu}_1(i_{m,n}, x)$$

for all $(i_{m,n}, x) \in \mathcal{W}$. We now get for almost every trajectory $(\underline{s}_1, \underline{s}_2, \dots) \in \mathcal{W}_\pi^{\mathbb{N}}$ of $(\widehat{\mathbf{Y}}_k)_{k\in\mathbb{N}}$:

$$
\begin{aligned}
&\lim_{k\to\infty} -\frac{1}{k}\log\mathbb{P}\big[\widehat{\mathbf{Y}}_1 = \underline{s}_1, \dots, \widehat{\mathbf{Y}}_k = \underline{s}_k\big] \\
={}& \lim_{k\to\infty} -\frac{1}{k}\log \sum_{\substack{\underline{y}_1, \dots, \underline{y}_{k+1}\in\mathcal{W}: \\ \pi(\underline{y}_j, \underline{y}_{j+1})=\underline{s}_j}} \mathbb{P}\big[(\hat{\mathbf{i}}_1, \mathbf{x}_1) = \underline{y}_1, \dots, (\hat{\mathbf{i}}_{k+1}, \mathbf{x}_{k+1}) = \underline{y}_{k+1}\big] \\
={}& \lim_{k\to\infty} -\frac{1}{k}\log \sum_{\substack{\underline{y}_1, \dots, \underline{y}_{k+1}\in\mathcal{W}: \\ \pi(\underline{y}_j, \underline{y}_{j+1})=\underline{s}_j}} \hat{\mu}_1(\underline{y}_1) \cdot \hat{q}(\underline{y}_1, \underline{y}_2) \cdot \ldots \cdot \hat{q}(\underline{y}_k, \underline{y}_{k+1}) \\
={}& \lim_{k\to\infty} -\frac{1}{k}\log \sum_{\substack{\underline{y}_1, \dots, \underline{y}_{k+1}\in\mathcal{W}: \\ \pi(\underline{y}_j, \underline{y}_{j+1})=\underline{s}_j}} \nu(\underline{y}_1)\hat{q}(\underline{y}_1, \underline{y}_2) \cdot \ldots \cdot \hat{q}(\underline{y}_k, \underline{y}_{k+1}) \\
={}& \lim_{k\to\infty} -\frac{1}{k}\log\mathbb{P}\big[\mathbf{Z}_1 = \underline{s}_1, \dots, \mathbf{Z}_k = \underline{s}_k\big] = H(\mathbf{Z}).
\end{aligned}
$$

$\square$

As a consequence we obtain the next statement:

Corollary 5.5.

$$\lim_{k\to\infty} -\frac{1}{k}\int \log\mathbb{P}\big[\widehat{\mathbf{Y}}_1 = \underline{s}_1, \dots, \widehat{\mathbf{Y}}_k = \underline{s}_k\big]\, d\mathbb{P}(\underline{s}_1, \underline{s}_2, \dots) = H(\mathbf{Z}).$$

Proof. Since $|\mathcal{W}| < \infty$ by construction, there is $\varepsilon_0 > 0$ such that

$$1 \geq \hat{q}(\underline{y}_1, \underline{y}_2) \geq \varepsilon_0 \quad \text{for all } \underline{y}_1, \underline{y}_2 \in \mathcal{W}.$$

Hence,

$$0 \leq -\frac{1}{k} \log \mathbb{P}\big[\widehat{\mathbf{Y}}_1 = \underline{s}_1, \ldots, \widehat{\mathbf{Y}}_k = \underline{s}_k\big] \leq -\frac{1}{k} \log c - \log \varepsilon_0,$$

where $c = \min\{\hat{\mu}_1(\underline{y}) \mid \underline{y} \in \mathcal{W}\}$. Therefore, we may exchange integral and limit, which yields the claim. $\square$

Let $x = x_1 \ldots x_n \in \mathcal{L}$, $n = |x| \geq 2$, be on the boundary of some cone C. Define

$$\hat{l}(x_1 \ldots x_n) := -\log \sum_{bc \in \mathcal{A}^2 : x_1 \ldots x_{n-2}bc \in \partial C} L(o, x_1 \ldots x_{n-2}bc).$$

Proposition 5.6.

$$\lim_{k \to \infty} \frac{\hat{l}(X_{\mathbf{e}_k})}{k} = H(\mathbf{Z}) \quad \text{almost surely.}$$

Proof. Assume that $\mathbf{W}_j = y_j a_j b_j$, where $y_j \in \mathcal{A}^*$ and $a_j b_j \in \mathcal{A}^2$ with $0 \leq j \leq k$. That is, $X_{\mathbf{e}_j} = y_1 y_2 \ldots y_j a_j b_j$. By definition, $X_{\mathbf{e}_j}$ is on the boundary of some cone, which we denote by C_j.

Let $j \in \mathcal{I}$. Recall that the covering of $\mathcal{L}$ consists of n_0 subcones. Each of these subcones $C_i^{(0)}$ has again a covering consisting of n_i subcones of type j. Write $N_j := \sum_{i=1}^{n_0} n_i$, and we denote by $C_{j,k}^{(1)}$ these different subcones with $1 \leq k \leq N_j$. Furthermore, we write $w \sim yab$ if $w = ycd \in \partial C(yab)$ for $y \in \mathcal{A}^*$ and $ab, cd \in \mathcal{A}^2$, that is, $w \sim yab$ if w lies on the same boundary of a cone as yab (namely the cone $C(yab)$).

Moreover, we have for all $j \in \mathcal{I}$ and $w_1 ab, w_2 ab \in \bigcup_{i=1}^{N_j} \partial C_{j,k}^{(1)}$ that $\mathbb{P}[X_{\mathbf{e}_1} = w_1 ab] > 0$ if and only if $\mathbb{P}[X_{\mathbf{e}_1} = w_2 ab] > 0$. Therefore, there are $c, C > 0$ such that

$$c \cdot \mathbb{P}[X_{\mathbf{e}_1} = w_2 ab] \leq \mathbb{P}[X_{\mathbf{e}_1} = w_1 ab] \leq C \cdot \mathbb{P}[X_{\mathbf{e}_1} = w_2 ab]$$

for all $w_1 ab, w_2 ab \in \bigcup_{i=1}^{N_j} \partial C_{j,k}^{(1)}$. Assume now that $\tau(C(a_1 b_1)) = j \in \mathcal{I}$. Observe that $\partial C(y_0 y_1 a_1 b_1) = \{y_0 y_1 c_1 d_1, \ldots, y_0 y_1 c_\kappa d_\kappa\}$ implies that $C_{j,k}^{(1)}$ has the form $\{wc_1 d_1, \ldots, wc_\kappa d_\kappa\}$ for some suitable $w \in \mathcal{A}^*$. We have:

$$N_j \cdot \sum_{\substack{w_1 \in \mathcal{L}: \\ w_1 \sim y_0 y_1 a_1 b_1}} \sum_{\substack{w_2, \ldots, w_k \in \mathcal{W}_0: \\ w_i \sim y_i a_i b_i}} \mathbb{P}[X_{\mathbf{e}_1} = w_1] q(y_1[w_1], w_2) \cdot q(w_2, w_3) \cdot \ldots \cdot q(w_{k-1}, w_k)$$

$$= \sum_{k=1}^{N_j} \sum_{\substack{w_1 \in \mathcal{L}: \\ w_1 \sim y_0 y_1 a_1 b_1}} \sum_{\substack{w_2, \ldots, w_k \in \mathcal{W}_0: \\ w_i \sim y_i a_i b_i}} \mathbb{P}[X_{\mathbf{e}_1} = w_1] q(y_1[w_1], w_2) \cdot q(w_2, w_3) \cdot \ldots \cdot q(w_{k-1}, w_k)$$

$$\leq C \cdot \sum_{k=1}^{N_j} \sum_{\substack{w_1 \in \partial C_{j,k}^{(1)}}} \sum_{\substack{w_2, \ldots, w_k \in \mathcal{W}_0: \\ w_i \sim y_i a_i b_i}} \mathbb{P}[X_{\mathbf{e}_1} = w_1] \frac{\xi([w_2])}{\xi([w_1])} \mathbb{L}(w_1, w_2) \cdot q(w_2, w_3) \cdot \ldots \cdot q(w_{k-1}, w_k)$$

$$= C \cdot \mathbb{P}\big[\mathbf{Y}_1 = (j, t^{(1)}), \ldots, \mathbf{Y}_{k-1} = (s_{k-1}, t^{(k-1)})\big], \tag{5.1}$$

where the values of $s_2, \ldots, s_{k-1}$ and $t^{(1)}, \ldots, t^{(k-1)}$ are determined by the values of $\mathbf{W}_j = y_j a_j b_j$. Analogously,

$$N_j \cdot \sum_{\substack{w_1 \in \mathcal{L}: \\ w_1 \sim y_0 y_1 a_1 b_1}} \sum_{\substack{w_2, \ldots, w_k \in \mathcal{W}_0: \\ w_i \sim y_i a_i b_i}} \mathbb{P}[X_{\mathbf{e}_1} = w_1] q(y_1[w_1], w_2) \cdot q(w_2, w_3) \cdot \ldots \cdot q(w_{k-1}, w_k)$$

$$\geq \quad c \cdot \mathbb{P}\big[\mathbf{Y}_1 = (j, t^{(1)}), \ldots, \mathbf{Y}_{k-1} = (s_{k-1}, t^{(k-1)})\big]. \tag{5.2}$$

Recall that $G(o, w) = G(o, o) L(o, w)$ for all $w \in \mathcal{L}$ and that $\xi(\cdot)$ can only take finitely many values. Writing $X_{\mathbf{e}_1} = x_1 \ldots x_n$ and $j = \tau(C(X_{\mathbf{e}_1}))$, we now can conclude as follows:

$$\lim_{k \to \infty} \frac{\hat{l}(X_{\mathbf{e}_k})}{k}$$

$$= \lim_{k \to \infty} -\frac{1}{k} \log \sum_{bc \in \mathcal{A}^2 : x_1 \ldots x_{n-2} bc \in \partial C_k} L(o, x_1 \ldots x_{k-2} bc)$$

$$= \lim_{k \to \infty} -\frac{1}{k} \log \left[\sum_{\substack{w_1 \in \mathcal{L}: \\ w_1 \sim y_0 y_1 a_1 b_1}} \sum_{\substack{w_2, \ldots, w_k \in \mathcal{W}_0: \\ w_i \sim y_i a_i b_i}} L(o, w_1) \mathbb{L}(w_1, w_2) \cdot \ldots \cdot \mathbb{L}(w_{k-1}, w_k) \right]$$

$$= \lim_{k \to \infty} -\frac{1}{k} \log \left[\sum_{\substack{w_1 \in \mathcal{L}: \\ w_1 \sim y_0 y_1 a_1 b_1}} \sum_{\substack{w_2, \ldots, w_k \in \mathcal{W}_0: \\ w_i \sim y_i a_i b_i}} G(o, w_1) \xi([w_1]) \cdot \frac{\xi([w_2])}{\xi([w_1])} \mathbb{L}(w_1, w_2) \cdot \ldots \cdot \right.$$

$$\left. \cdot \frac{\xi([w_k])}{\xi([w_{k-1}])} \mathbb{L}(w_{k-1}, w_k) \right]$$

$$= \lim_{k \to \infty} -\frac{1}{k} \log \left[\sum_{\substack{w_1 \in \mathcal{L}: \\ w_1 \sim y_0 y_1 a_1 b_1}} \sum_{\substack{w_2, \ldots, w_k \in \mathcal{W}_0: \\ w_i \sim y_i a_i b_i}} \mathbb{P}[X_{\mathbf{e}_1} = w_1] q(y_1[w_1], w_2) \cdot \ldots \cdot q(w_{k-1}, w_k) \right]$$

$$= \lim_{k \to \infty} -\frac{1}{k} \log \left[N_j \sum_{\substack{w_1 \in \mathcal{L}: \\ w_1 \sim y_0 y_1 a_1 b_1}} \sum_{\substack{w_2, \ldots, w_k \in \mathcal{W}_0: \\ w_i \sim y_i a_i b_i}} \mathbb{P}[X_{\mathbf{e}_1} = w_1] q(y_1[w_1], w_2) \cdot \ldots \cdot q(w_{k-1}, w_k) \right]$$

$$= \lim_{k \to \infty} -\frac{1}{k} \log \mathbb{P}\big[\mathbf{Y}_1 = (j, t^{(1)}), \ldots, \mathbf{Y}_{k-1} = (s_{k-1}, t^{(k-1)}\big] = H(\mathbf{Z}).$$

The last equation follows from (5.1) and (5.2). We need those important estimates since the first coordinate of $\mathbf{Y}_1$ describes only the cone type of $X_{\mathbf{e}_1}$ but there may be several cones of the same type $j = \tau(C(X_{\mathbf{e}_1}))$.

$\square$

Recall the definition of $l(x_1 \ldots x_n) = -\log L(o, x_1 \ldots x_n)$ for $x = x_1 \ldots x_n \in \mathcal{L}$.

Corollary 5.7.

$$\lim_{n \to \infty} \frac{l(X_{\mathbf{e}_k})}{k} = H(\mathbf{Z}) \quad \text{almost surely.}$$

Proof. It suffices to compare $\hat{l}(X_{\mathbf{e}_k})$ with $l(X_{\mathbf{e}_k})$. Assume for a moment that $X_{\mathbf{e}_k} = x_1 \ldots x_n$ and that $X_{\mathbf{e}_k}$ is on the boundary of the cone C. Then, the probability of walking *inside*

C from $x_1 \ldots x_n \in \partial C$ to any $x_1 \ldots x_{n-2}ab \in \partial C$ can be bounded from below by some constant ε_0, because the probabilities depend only on $x_{n-1}x_n$ and $ab \in \mathcal{A}^2$. Therefore,

$$L(o, X_{\mathbf{e}_k}) \cdot \varepsilon_0 \quad \leq \sum_{bc \in \mathcal{A}^2 : x_1 \ldots x_{n-2}ab \in \partial C} L(x_1 \ldots x_{m-2}ab),$$

$$\sum_{ab \in \mathcal{A}^2 : x_1 \ldots x_{n-2}ab \in \partial C} L(o, x_1 \ldots x_{n-2}ab) \cdot \varepsilon_0 \quad \leq \quad |\mathcal{A}^2| \cdot L(o, X_{\mathbf{e}_k}).$$

Taking logarithms, dividing by k and letting k tend to infinity yields the claim. $\qquad\square$

Now we come to an important law of large numbers. For this purpose, define $d(x, y) = |y| - |x|$ for $x, y \in \mathcal{L}$ with $|x| \leq |y|$, where $|\cdot|$ is the natural word length. Denote by ν_0 the invariant probabilty measure of $(\mathbf{W}_k)_{k \in \mathbb{N}}$ and define

$$\lambda := \sum_{x,y \in \mathcal{W}_0} \nu_0(x) \cdot q(x, y) \cdot d(x, y). \tag{5.3}$$

Then:

Proposition 5.8.
$$\lim_{k \to \infty} \frac{l(X_n)}{n} = \ell \cdot \lambda^{-1} \cdot H(\mathbf{Z}) \quad \text{almost surely.}$$

Proof. Define
$$\hat{\mathbf{e}}_k := \sup\{m \in \mathbb{N} \big| |X_m| = k\}.$$
Transience yields $\hat{\mathbf{e}}_k < \infty$ almost surely for all $k \in \mathbb{N}$. Define the maximal exit times at time $n \in \mathbb{N}$ as

$$\begin{aligned} \mathbf{k}(n) &:= \max\{k \in \mathbb{N} \mid \hat{\mathbf{e}}_k \leq n\}, \\ \mathbf{t}(n) &:= \max\{k \in \mathbb{N} \mid \mathbf{e}_k \leq n\}. \end{aligned}$$

Obviously, $\mathbf{k}(n) \geq \mathbf{t}(n)$ and each exit time $\mathbf{e}_k$ corresponds to exactly one $\hat{\mathbf{e}}_l$ with $l \geq k$. First, we rewrite

$$\frac{l(X_n)}{n} = \frac{l(X_n) - l(X_{\mathbf{e}_{\mathbf{t}(n)}})}{n} + \frac{l(X_{\mathbf{e}_{\mathbf{t}(n)}})}{\mathbf{t}(n)} \cdot \frac{\mathbf{t}(n)}{\mathbf{k}(n)} \cdot \frac{\mathbf{k}(n)}{\hat{\mathbf{e}}_{\mathbf{k}(n)}} \cdot \frac{\hat{\mathbf{e}}_{\mathbf{k}(n)}}{n}. \tag{5.4}$$

Let ε_1 be the minimal occuring positive single-step transition probability. Since the sub-cones of coverings of bigger cones are nested at bounded distance we have $\hat{\mathbf{e}}_{\mathbf{k}(n)} \geq \mathbf{e}_{\mathbf{t}(n)} \geq \hat{\mathbf{e}}_{\mathbf{k}(n)-D}$ for some suitable $D \in \mathbb{N}$. The first quotient on the right hand side of (5.4) tends to zero since

$$\begin{aligned} L(o, X_n) \cdot \varepsilon_1^{n - \mathbf{e}_{\mathbf{t}(n)}} &\leq L(o, X_{\mathbf{e}_{\mathbf{t}(n)}}) \quad \text{(due to weak symmetry),} \\ L(o, X_{\mathbf{e}_{\mathbf{t}(n)}}) \varepsilon_1^{n - \mathbf{e}_{\mathbf{t}(n)}} &\leq L(o, X_n) \end{aligned}$$

and due to (follows completely analogously as in [16, Proof of Theorem D])

$$\frac{\mathbf{e}_{\mathbf{t}(n)}}{n} \leq \frac{\hat{\mathbf{e}}_{\mathbf{k}(n)}}{n} \xrightarrow{n \to \infty} 1, \quad \frac{\mathbf{e}_{\mathbf{t}(n)}}{n} \geq \frac{\hat{\mathbf{e}}_{\mathbf{k}(n)-D}}{n} \xrightarrow{n \to \infty} 1,$$

which in turn yields $(n - \mathbf{e}_{\mathbf{t}(n)})/n \to 0$ as $n \to \infty$.

By Corollary 5.7, $l(X_{\mathbf{e}_{\mathbf{t}(n)}})/\mathbf{t}(n)$ tends to $H(\mathbf{Z})$. On the other hand side, $\hat{\mathbf{e}}_k/k$ tends almost surely to $1/\ell$ and $\hat{\mathbf{e}}_{\mathbf{k}(n)}/n$ tends to 1 almost surely; see [6, Proposition 2.3]. It remains to prove that the limit $\lim_{k\to\infty} \mathbf{k}(n)/\mathbf{t}(n)$ exists. Clearly,

$$\frac{\mathbf{k}(n)}{\mathbf{t}(n)} = \frac{1}{\mathbf{t}(n)}\Big(d(o, X_{\mathbf{e}_{\mathbf{t}_1}}) + \sum_{i=1}^{\mathbf{t}(n)-1} d(X_{\mathbf{e}_i}, X_{\mathbf{e}_{i+1}}) + d(X_{\mathbf{e}_{\mathbf{t}(n)}}, X_{\hat{\mathbf{e}}_{\mathbf{k}(n)}})\Big).$$

Note that

$$d(o, X_{\mathbf{e}_{\mathbf{t}_1}}) \leq D_1, \quad d(X_{\mathbf{e}_{\mathbf{t}(n)}}, X_{\hat{\mathbf{e}}_{\mathbf{k}(n)}}) \leq D_2$$

for a suitable constants D_1 and D_2. Thus, it is sufficient to consider

$$\frac{1}{k}\sum_{i=1}^{k} d(X_{\mathbf{e}_i}, X_{\mathbf{e}_{i+1}}).$$

Since $d(X_{\mathbf{e}_i}, X_{\mathbf{e}_{i+1}})$ can be computed from $\mathbf{W}_i$ and $\mathbf{W}_{i+1}$, we may apply the ergodic theorem for positive recurrent Markov chains on the process $\big((\mathbf{W}_j, \mathbf{W}_{j+1})\big)_{j\in\mathbb{N}}$ which yields almost surely

$$\lim_{k\to\infty} \frac{1}{k}\sum_{i=1}^{k} d(X_{\mathbf{e}_i}, X_{\mathbf{e}_{i+1}}) = \sum_{x,y\in\mathcal{W}_0} \nu_0(x)q(x,y)d(x,y) = \lambda.$$

This finishes the proof and gives the proposed formula. $\qquad\square$

6. Existence of Entropy

We follow the reasoning of [7] for the proof of existence of the entropy. First, we need the following lemma:

Lemma 6.1. *There is $R > 1$ such that $G(o, w|R) < \infty$ for all $w \in \mathcal{L}$.*

Proof. A simple adaption of the proof of [12, Proposition 8.2] shows that $G(v, w|z)$ has radius of convergence $R(v, w) > 1$. At this point we need suffix-irreducibility. With the help of this fact we are able to prove the lemma in several steps:

(1) There is $R_0 > 1$ such that $L(o, abc|R_0) < \infty$ for all $abc \in \mathcal{A}^3$: this follows from the inequality $G(o, w|z) \geq L(o, w|z)$.

(2) There is $R_1 > 1$ such that $\bar{G}(ab, cd|R_1) < \infty$ for all $ab, cd \in \mathcal{A}^2$: this follows from the inequality $G(ab, cd|z) \geq \bar{G}(ab, cd|z)$.

(3) Since for $a, b, c, d, e \in \mathcal{A}$

$$\bar{L}(ab, cde|z) = \sum_{d_1 e_1 \in \mathcal{A}^2} p(ab, cd_1 e_1) \cdot z \cdot \bar{G}(d_1 e_1, de|z),$$

we have $\bar{L}(ab, cde|R_1) < \infty$.

(4) By $G(o, w|z) = G(o, o|z)L(o, w|z)$ and Equation (3.4), we get $G(o, w|R) < \infty$ for all $w \in \mathcal{L}$, where $R = \min\{R(o, o), R_0, R_1\} > 1$.

This finishes the proof. $\qquad\square$

In the following let $\varrho \in [1, R)$.

Lemma 6.2. *There are constants D_1 and $D_2 > 0$ such that for all $m, n \in \mathbb{N}_0$*

$$p^{(m)}(o, X_n) \leq D_1 \cdot D_2^n \cdot \varrho^{-m}.$$

Proof. Denote by $\mathcal{C}_\varrho$ the circle with radius ϱ in the complex plane centered at 0. A straight-forward computation shows for $m \in \mathbb{N}_0$:

$$\frac{1}{2\pi i} \oint_{\mathcal{C}_\varrho} z^m \frac{dz}{z} = \begin{cases} 1, & \text{if } m = 0, \\ 0, & \text{if } m \neq 0. \end{cases}$$

Let $w = w_1 \ldots w_t \in \mathcal{L}$. An application of Fubini's Theorem yields

$$\frac{1}{2\pi i} \oint_{\mathcal{C}_\varrho} G(o, w|z)\, z^{-m} \frac{dz}{z} = \frac{1}{2\pi i} \oint_{\mathcal{C}_\varrho} \sum_{k \geq 0} p^{(k)}(o, w) z^k\, z^{-m} \frac{dz}{z}$$

$$= \frac{1}{2\pi i} \sum_{k \geq 0} p^{(k)}(o, w) \oint_{\mathcal{C}_\varrho} z^{k-m} \frac{dz}{z} = p^{(m)}(o, w).$$

Since $G(o, w|z)$ is analytic on $\mathcal{C}_\varrho$, we have $|G(o, w|z)| \leq G(o, w|\varrho)$ for all $|z| = \varrho$. Thus,

$$p^{(m)}(o, w) \leq \frac{1}{2\pi} \cdot \varrho^{-m-1} \cdot G(o, w|\varrho) \cdot 2\pi \varrho = G(o, w|\varrho) \cdot \varrho^{-m}.$$

Set $L := 1 \vee \max\{\bar{L}(ab, cde|\varrho) \mid a, b, c, d, e \in \mathcal{A}\}$ and $D_0 := G(o, o|\varrho) \cdot \sum_{abc \in \mathcal{A}^3} L(o, abc|\varrho)$. An application of Equation (3.4) provides for $t \geq 3$

$$G(o, w|\varrho) = G(o, o|\varrho) \cdot L(o, w_1 \ldots w_t) \leq D_0 \cdot |\mathcal{A}|^{2(t-3)} \cdot L^{t-3}.$$

Set $D_1 := D_0 \vee \max\{G(o, w|\varrho) | w \in \mathcal{L}, |w| \leq 2\}$. Since $|X_n| \leq n$, we obtain by setting $D_2 := |\mathcal{A}|^2 \cdot L$

$$p^{(m)}(o, X_n) \leq D_1 \cdot |\mathcal{A}|^{2t} L^t \cdot \varrho^{-m} \leq D_1 \cdot |\mathcal{A}|^{2n} L^n \cdot \varrho^{-m} = D_1 \cdot D_2^n \cdot \varrho^{-m}.$$

$\square$

The following technical lemma will be used in the proof of the next theorem:

Lemma 6.3. *Let $(A_n)_{n \in \mathbb{N}}$, $(a_n)_{n \in \mathbb{N}}$, $(b_n)_{n \in \mathbb{N}}$ be sequences of strictly positive numbers with $A_n = a_n + b_n$. Assume that $\lim_{n \to \infty} -\frac{1}{n} \log A_n = c \in [0, \infty)$ and that $\lim_{n \to \infty} b_n / q^n = 0$ for all $q \in (0, 1)$. Then $\lim_{n \to \infty} -\frac{1}{n} \log a_n = c$.*

Proof. A proof can be found in [7, Lemma 3.5]. $\square$

Lemma 6.4. *For $n \in \mathbb{N}$, consider the function $f_n : \mathcal{L} \to \mathbb{R}$ defined by*

$$f_n(w) := \begin{cases} -\frac{1}{n} \log \sum_{m=0}^{n^2} p^{(m)}(o, w), & \text{if } p^{(n)}(o, w) > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Then there are constants d and D such that $d \leq f_n(w) \leq D$ for all $n \in \mathbb{N}$ and $w \in \mathcal{L}$.

Proof. Let $w \in \mathcal{L}$ and $n \in \mathbb{N}$ with $p^{(n)}(o, w) > 0$. Denote by R the radius of convergence of $G(w, w|z)$. By Inequality (3.1), we get

$$\sum_{m=0}^{n^2} p^{(m)}(o, w) \leq G(o, w|1) = F(o, w|1) \cdot G(w, w|1) \leq \frac{1}{1 - \frac{1}{R}},$$

that is,

$$f_n(w) \geq -\frac{1}{n} \log \frac{1}{1 - \frac{1}{R}} \geq -\log \frac{1}{1 - \frac{1}{R}}.$$

For the upper bound, observe that $w \in \mathcal{L}$ with $p^{(n)}(o, w) > 0$ can be reached from o in n steps with a probability of at least ε_0^n, where

$$\varepsilon_0 := \min\{p(w_1, w_2) \mid w_1, w_2 \in \mathcal{A}^*, p(w_1, w_2) > 0\} > 0$$

is independent from w. Thus, the sum $\sum_{m=0}^{n^2} p^{(m)}(o, w)$ has a value greater or equal to ε_0^n. Hence, $f_n(x) \leq -\log \varepsilon_0$. $\qquad\square$

Now we can finally prove:

Theorem 6.5. *The asymptotic entropy h of $(X_n)_{n \in \mathbb{N}_0}$ exists and equals $h = \ell \cdot \lambda^{-1} \cdot H(\mathbf{Z})$.*

Proof. We can rewrite $\ell \cdot \lambda^{-1} \cdot H(\mathbf{Z})$ as

$$
\begin{aligned}
\ell \cdot \lambda^{-1} \cdot H(\mathbf{Z}) &= \int \ell \cdot \lambda^{-1} \cdot H(\mathbf{Z}) \, d\mathbb{P} = \int \lim_{n \to \infty} -\frac{1}{n} l(X_n) \, d\mathbb{P} \\
&= \int \lim_{n \to \infty} -\frac{1}{n} \log L\big(o, X_n(\omega)\big|1\big) \, d\mathbb{P}(\omega) \\
&= \int \lim_{n \to \infty} -\frac{1}{n} \log \frac{G\big(o, X_n(\omega)|1\big)}{G(o, o|1)} \, d\mathbb{P}(\omega) = \int \lim_{n \to \infty} -\frac{1}{n} \log G\big(o, X_n(\omega)|1\big) \, d\mathbb{P}(\omega).
\end{aligned}
$$

Since

$$G(o, X_n|1) = \sum_{m \geq 0} p^{(m)}(o, X_n) \geq p^{(n)}(o, X_n) = \pi_n(X_n),$$

we have

$$\ell \cdot \lambda^{-1} \cdot H(\mathbf{Z}) \leq \int \liminf_{n \to \infty} -\frac{1}{n} \log \pi_n\big(X_n(\omega)\big) \, d\mathbb{P}(\omega). \tag{6.1}$$

The next aim is to prove $\limsup_{n \to \infty} -\frac{1}{n} \mathbb{E}\big[\log \pi_n(X_n)\big] \leq h$. We now apply Lemma 6.3 by setting

$$A_n := \sum_{m \geq 0} p^{(m)}(o, X_n), \ a_n := \sum_{m=0}^{n^2} p^{(m)}(o, X_n) \text{ and } b_n := \sum_{m \geq n^2 + 1} p^{(m)}(o, X_n).$$

By Lemma 6.2,

$$b_n \leq \sum_{m \geq n^2 + 1} D_1 \cdot D_2^n \cdot \varrho^{-m} = D_1 \cdot D_2^n \cdot \frac{\varrho^{-n^2 - 1}}{1 - \varrho^{-1}}.$$

Therefore, b_n decays faster than any geometric sequence. Applying Lemma 6.3 yields

$$\ell \cdot \lambda^{-1} \cdot h(\mathbf{Z}) = \lim_{n \to \infty} -\frac{1}{n} \log \sum_{m=0}^{n^2} p^{(m)}(o, X_n).$$

By Lemma 6.4, we may apply the Dominated Convergence Theorem and get:

$$
\begin{aligned}
\ell \cdot \lambda^{-1} \cdot H(\mathbf{Z}) = \int \ell \cdot \lambda^{-1} \cdot H(\mathbf{Z}) \, d\mathbb{P} \;&=\; \int \lim_{n \to \infty} -\frac{1}{n} \log \sum_{m=0}^{n^2} p^{(m)}(o, X_n) \, d\mathbb{P} \\
&=\; \lim_{n \to \infty} \int -\frac{1}{n} \log \sum_{m=0}^{n^2} p^{(m)}(o, X_n) \, d\mathbb{P} \\
&=\; \lim_{n \to \infty} -\frac{1}{n} \sum_{w \in \mathcal{L}} p^{(n)}(o, w) \log \sum_{m=0}^{n^2} p^{(m)}(o, w).
\end{aligned}
$$

For $w \in \mathcal{L}$, set

$$\mu_0(w) = \frac{1}{n^2 + 1} \sum_{m=0}^{n^2} p^{(m)}(o, w).$$

Recall that *Shannon's Inequality* gives

$$-\sum_{w \in \mathcal{L}} p^{(n)}(o, w) \log \mu(w) \geq -\sum_{w \in \mathcal{L}} p^{(n)}(o, w) \log p^{(n)}(o, w)$$

for every finitely supported probability measure μ on $\mathcal{L}$. We apply now this inequality on μ_0:

$$
\begin{aligned}
\ell \cdot \lambda^{-1} \cdot H(\mathbf{Z}) \;&\geq\; \limsup_{n \to \infty} \frac{1}{n} \sum_{w \in \mathcal{L}} p^{(n)}(o, w) \log(n^2 + 1) - \frac{1}{n} \sum_{w \in \mathcal{L}} p^{(n)}(o, w) \log p^{(n)}(o, w) \\
&=\; \limsup_{n \to \infty} -\frac{1}{n} \int \log \pi_n(X_n) \, d\mathbb{P}.
\end{aligned}
$$

Now we can conclude with Fatou's Lemma:

$$
\begin{aligned}
h = \ell \cdot \lambda^{-1} \cdot H(\mathbf{Z}) \;&\leq\; \int \liminf_{n \to \infty} \frac{-\log \pi_n(X_n)}{n} d\mathbb{P} \leq \liminf_{n \to \infty} \int \frac{-\log \pi_n(X_n)}{n} d\mathbb{P} \\
&\leq\; \limsup_{n \to \infty} \int \frac{-\log \pi_n(X_n)}{n} d\mathbb{P} \leq \ell \cdot \lambda^{-1} \cdot H(\mathbf{Z}) = h. \qquad (6.2)
\end{aligned}
$$

Thus, $h = \lim_{n \to \infty} -\frac{1}{n} \mathbb{E}\big[\log \pi_n(X_n)\big]$ exists and the limit equals $\ell \cdot \lambda^{-1} \cdot H(\mathbf{Z})$. $\qquad \square$

We get the following types of convergence:

Corollary 6.6.   (1) *For almost every path of the random walk* $(X_n)_{n \in \mathbb{N}_0}$,

$$h = \liminf_{n \to \infty} -\frac{\log \pi_n(X_n)}{n}.$$

(2) *Convergence in probability:*

$$-\frac{1}{n}\log\pi_n(X_n) \xrightarrow{\mathbb{P}} h.$$

(3) *Convergence in L_1:*

$$-\frac{1}{n}\log\pi_n(X_n) \xrightarrow{L_1} h.$$

Proof. The proofs are completely analogous to the proofs in [7, Corollary 3.9, Lemma 3.10], where [7, Lemma 3.10] holds also in the case $h = 0$. $\qquad\square$

Corollary 6.7. *The entropy is the rate of escape with respect to the Greenian distance, that is,*

$$h = \lim_{n\to\infty} -\frac{1}{n}\log G(o, X_n|1)\, d\mathbb{P}.$$

Proof. This follows from the simple fact $G(o, X_n|1) = G(o, o|1)L(o, X_n|1)$ and Proposition 5.8. $\qquad\square$

7. Calculation of the Entropy

In order to compute $h = \ell \cdot \lambda^{-1} \cdot H(\mathbf{Z})$ we have to calculate the three factors: while there is a formula for ℓ (given in [6, Theorem 2.4]) and there is also a formula for λ (given in (5.3)), it remains to explain how to calculate $H(\mathbf{Z})$. For this purpose, define

$$H(\mathbf{Z}_1,\ldots,\mathbf{Z}_n) := -\sum_{\underline{s}_1,\ldots,\underline{s}_n\in\mathcal{W}_\pi} \mathbb{P}\big[\mathbf{Z}_1 = \underline{s}_1,\ldots,\mathbf{Z}_n = \underline{s}_n\big]\log\mathbb{P}\big[\mathbf{Z}_1 = \underline{s}_1,\ldots,\mathbf{Z}_n = \underline{s}_n\big],$$

and let the *conditional entropy* $H(\mathbf{Z}_n|\mathbf{Z}_1,\ldots,\mathbf{Z}_{n-1})$ be defined as

$$-\sum_{\underline{s}_1,\ldots,\underline{s}_n\in\mathcal{W}_\pi} \mathbb{P}\big[\mathbf{Z}_1 = \underline{s}_1,\ldots,\mathbf{Z}_n = \underline{s}_n\big]\log\mathbb{P}\big[\mathbf{Z}_n = \underline{s}_n\big|\mathbf{Z}_1 = \underline{s}_1,\ldots,\mathbf{Z}_{n-1} = \underline{s}_{n-1}\big].$$

By [4, Theorem 4.2.1], $H(\mathbf{Z}) = \lim_{n\to\infty}\frac{1}{n}H(\mathbf{Z}_1,\ldots,\mathbf{Z}_n)$. In general, the computation of $H(\mathbf{Z})$ is a hard task. But there is a simple way in order to calculate $H(\mathbf{Z})$ numerically, which is due to the inequalities

$$H\big(\mathbf{Z}_n\big|((\hat{\mathbf{i}}_1^{(\nu)}, \mathbf{x}_1^{(\nu)}), (\hat{\mathbf{i}}_2^{(\nu)}, \mathbf{x}_2^{(\nu)})), \mathbf{Z}_1,\ldots,\mathbf{Z}_{n-1}\big) \leq H(\mathbf{Z}) \leq H(\mathbf{Z}_n\mid\mathbf{Z}_1,\ldots,\mathbf{Z}_{n-1}) \qquad (7.1)$$

for all $n\in\mathbb{N}$; see [4, Theorem 4.5.1]. In particular, it is even shown that

$$H(\mathbf{Z}_n\mid\mathbf{Z}_1,\ldots,\mathbf{Z}_{n-1}) - H\big(\mathbf{Z}_n\big|((\hat{\mathbf{i}}_1^{(\nu)}, \mathbf{x}_1^{(\nu)}), (\hat{\mathbf{i}}_2^{(\nu)}, \mathbf{x}_2^{(\nu)})), \mathbf{Z}_1,\ldots,\mathbf{Z}_{n-1}\big) \xrightarrow{n\to\infty} 0.$$

Hence, one can calculate $H(\mathbf{Z})$ numerically up to an arbitrarily small error. Furthermore:

Corollary 7.1. *If the random walk is expanding, then $h > 0$. Otherwise, $h = 0$.*

Proof. In the expanding case, the random walk has at least two possibilities for entering a subcone decsribed by $X_{\mathbf{e}_3}$ for every given value of $X_{\mathbf{e}_2}$. Thus,

$$H(\mathbf{Z}) \geq H\big(\mathbf{Z}_2\big|((\hat{\mathbf{i}}_1^{(\nu)}, \mathbf{x}_1^{(\nu)}), (\hat{\mathbf{i}}_2^{(\nu)}, \mathbf{x}_2^{(\nu)})), \mathbf{Z}_1\big) > 0,$$

which yields $h > 0$ due to (7.1). On the other hand side, if the random walk on $\mathcal{L}$ is not expanding, then each cone has a covering consisting of of only one single cone. Then the projections $\mathbf{Z}_n$ become deterministic and this implies

$$0 \leq H(\mathbf{Z}) \leq H(\mathbf{Z}_n \mid \mathbf{Z}_1, \ldots, \mathbf{Z}_{n-1}) = H(\mathbf{Z}_n) = 0.$$

$\square$

We call $ab \in \mathcal{A}^2$ *unambiguous* if $\partial C(ab) = \{ab\}$. In other words, whenever the random walk enters a subcone of type $C(wab)$, $w \in \mathcal{A}^*$, it must enter it through its single boundary point wab. This allows us to "cut" the random walk into pieces and to obtain another formula for the entropy $H(\mathbf{Z})$. For $n \in \mathbb{N}$, $x_2, \ldots, x_n \in \mathcal{W}_0$ and $ab \in \mathcal{A}^2$ define

$$w(ab, x_2, \ldots, x_n, x) \quad := \quad \mathbb{P}\big[\mathbf{W}_2 = x_2, \ldots, \mathbf{W}_n = x, [\mathbf{W}_n] = ab \big| \mathbf{W}_1 = ab\big],$$

$$\tilde{w}(ab, x_2, \ldots, x_n) \quad := \quad \sum_{\substack{y_2, \ldots, y_n \in \mathcal{W}_0: \\ y_i \sim x_i}} \mathbb{P}\big[\mathbf{W}_2 = y_2, \ldots, \mathbf{W}_n = y_n, [\mathbf{W}_n] = ab \big| \mathbf{W}_1 = ab\big],$$

where $\sim$ is the relation introduced in the proof of Proposition 5.6. In particular, $\tilde{y}(ab, x_2) = \mathbb{P}\big[\mathbf{W}_2 = x_2, [\mathbf{W}_2] = ab \big| \mathbf{W}_1 = ab\big]$. Denote by ν_1 the invariant probability measure of the process $(\mathbf{i}_k, \mathbf{W}_k)_{k \in \mathbb{N}}$ and set, for unambiguous $ab \in \mathcal{A}^2$,

$$\nu_{\mathbf{W}}(ab) := \sum_{(i_{m,n}, x) \in \mathcal{W}: [x] = ab} \nu_1(i_{m,n}, x).$$

Then:

Proposition 7.2. *If* $ab \in \mathcal{A}^2$ *is unambiguous, then*

$$h(\mathbf{Z}) = -\nu_{\mathbf{W}}(ab) \sum_{n \geq 1} \sum_{\substack{x_2, \ldots x_{n-1} \in \mathcal{W}_0: \ x_n \in \mathcal{W}_0: [x_n] = ab \\ [x_i] \neq ab}} \sum w(ab, x_2, \ldots, x_n) \log \tilde{w}(ab, x_2, \ldots, x_n).$$

Proof. By Propositions 5.3 and 5.4, we have that

$$-\frac{1}{n} \log \mathbb{P}[\mathbf{Y}_1 = \underline{s}_1, \ldots, \mathbf{Y}_n = \underline{s}_n] \xrightarrow{n \to \infty} H(\mathbf{Z})$$

for almost every realisation $(\underline{s}_1, \underline{s}_2, \ldots) \in \mathcal{W}_\pi^{\mathbb{N}}$. Observe that $\tau(\mathbf{W}_{n+1}) = ab$ is equivalent to $\mathbf{Y}_n = (t_n, \alpha_{t_n, m})$ for some cone type $t_n \in \mathcal{I}$, where α denotes the cone type of $C(ab)$ and $1 \leq m \leq n(t_n, \alpha)$. For any such trajectory, we define

$$N_0 := \min\big\{m \in \mathbb{N}\big|\tau(\mathbf{W}_{m+1}) = \alpha\big\} \text{ and } N_k := \min\big\{m \in \mathbb{N}\big|m > N_{k-1}, \tau(\mathbf{W}_{m+1}) = \alpha\big\}.$$

For any realisation $(\underline{s}_1, \underline{s}_2, \ldots) \in \mathcal{W}_\pi^{\mathbb{N}}$ and $n \in \mathbb{N}$, denote by $d(n)$ the maximal index k with $N_k \leq n$. Since $[\mathbf{W}_{N_k+1}] = ab$ for all $k \in \mathbb{N}$ we can use the strong Markov property as follows when $N_j < n$:

$$\mathbb{P}\big[\mathbf{Y}_{N_j+1} = \underline{s}_{N_j+1}, \ldots, \mathbf{Y}_n = \underline{s}_n \mid \mathbf{Y}_1 = \underline{s}_1, \ldots, \mathbf{Y}_{N_j} = \underline{s}_{N_j}\big]$$

$$= \quad \mathbb{P}\big[\mathbf{Y}_{N_j+1} = \underline{s}_{N_j+1}, \ldots, \mathbf{Y}_n = \underline{s}_n \mid [\mathbf{W}_{N_j+1}] = ab\big]$$

$$= \quad \mathbb{P}\big[\mathbf{Y}_{N_j+1} = \underline{s}_{N_j+1}, \ldots, \mathbf{Y}_n = \underline{s}_n \mid \mathbf{Y}_{N_j} = \underline{s}_{N_j}\big].$$

Therefore, we can rewrite the following probability:

$$\mathbb{P}\big[\mathbf{Y}_1 = \underline{s}_1, \ldots, \mathbf{Y}_{d(n)} = \underline{s}_{d(n)}\big]$$

$$= \mathbb{P}\big[\mathbf{Y}_1 = \underline{s}_1, \ldots, \mathbf{Y}_{N_0} = \underline{s}_{N_0}\big]\mathbb{P}\big[\mathbf{Y}_{N_0+1} = \underline{s}_{N_0+1}, \ldots, \mathbf{Y}_{N_1} = \underline{s}_{N_1} \mid \mathbf{Y}_{N_0} = \underline{s}_{N_0}\big] \cdot \ldots \cdot$$

$$\cdot \mathbb{P}\big[\mathbf{Y}_{N_{d(n)-1}+1} = \underline{s}_{N_{d(n)-1}+1}, \ldots, \mathbf{Y}_{N_{d(n)}} = \underline{s}_{N_{d(n)}} \mid \mathbf{Y}_{N_{d(n)-1}} = \underline{s}_{N_{d(n)-1}}\big].$$

Obviously, $d(n)/n$ tends almost surely to $\nu_{\mathbf{W}}(ab)$. Hence, if we consider only the subsequence where n equals one of the N_k's we obtain the following almost sure convergence:

$$-\frac{1}{n}\log\mathbb{P}\big[\mathbf{Y}_1 = \underline{s}_1, \ldots, \mathbf{Y}_{d(n)} = \underline{s}_{d(n)}\big]$$

$$= \quad -\frac{d(n)}{n}\frac{1}{d(n)}\Big[\log\mathbb{P}\big[\mathbf{Y}_1 = \underline{s}_1, \ldots, \mathbf{Y}_{N_0} = \underline{s}_{N_0}\big]$$

$$+ \log\mathbb{P}\big[\mathbf{Y}_{N_0+1} = \underline{s}_{N_0+1}, \ldots, \mathbf{Y}_{N_1} = \underline{s}_{N_1}\big|\mathbf{Y}_{N_0} = \underline{s}_{N_0}\big] + \ldots +$$

$$+ \log\mathbb{P}\big[\mathbf{Y}_{N_{d(n)-1}+1} = \underline{s}_{N_{d(n)-1}+1}, \ldots, \mathbf{Y}_{N_{d(n)}} = \underline{s}_{N_{d(n)}}\big|\mathbf{Y}_{N_{d(n)-1}} = \underline{s}_{N_{d(n)-1}}\big]\Big]$$

$$\xrightarrow{n\to\infty} \quad -\nu_{\mathbf{W}}(ab)\sum_{k\geq 1}\sum_{\substack{x_2,\ldots,x_{k-1}\in\mathcal{W}_0:\\ [x_i]\neq ab}}\sum_{\substack{x\in\mathcal{W}_0:\\ [x]=ab}} w(ab, x_2, \ldots, x_{k-1}, x)\log\tilde{w}(ab, x_2, \ldots, x_{k-1}, x).$$

This proves the claim. $\qquad\square$

We now state an inequality which connects entropy, drift and growth. For this purpose, define the *growth* of $\mathcal{A}^*$ as $g := \log|\mathcal{A}|$. Then we get:

Proposition 7.3. $h \leq \ell \cdot g$.

Proof. Let $\varepsilon > 0$. By Corollary 6.6 (1), there is some $N_\varepsilon \in \mathbb{N}$ such that for all $n \geq N_\varepsilon$:

$$1 - \varepsilon \leq \mathbb{P}\big[-\log\pi_n(X_n) \geq (h-\varepsilon)n, |X_n| \leq (\ell+\varepsilon)n\big] \leq e^{-(h-\varepsilon)n}\cdot|\mathcal{A}|^{(\ell+\varepsilon)n}.$$

Taking logarithms and dividing by n gives

$$(h - \varepsilon) + \frac{1}{n}\log(1-\varepsilon) \leq (\ell+\varepsilon)\cdot\log|\mathcal{A}|.$$

Making ε arbirtraily small yields the proposed claim. $\qquad\square$

Finally, we remark that the entropy is zero for recurrent random walks:

Corollary 7.4. *If $(X_n)_{n\in\mathbb{N}_0}$ is recurrent then $h = 0$.*

Proof. Clearly, $-\frac{1}{n}\mathbb{E}\big[\log\pi_n(X_n)\big] \geq 0$. Assume now that $\limsup_{n\to\infty} -\frac{1}{n}\mathbb{E}\big[\log\pi_n(X_n)\big] = c > 0$. Then there is a deterministic sequence $(n_k)_{k\in\mathbb{N}}$ such that, for any small $\varepsilon_1 > 0$,

$$-\frac{1}{n_k}\mathbb{E}\big[\log\pi_{n_k}(X_{n_k})\big] \geq c - \varepsilon_1 > 0 \qquad (7.2)$$

for all sufficiently large k. Denote by p_0 the minimal occuring positive single-step transition probability. Then $-\frac{1}{n_k}\log\pi_{n_k}(X_{n_k}) \leq -\log p_0$. Moreover, choose $N \in \mathbb{N}$ with $1/N < c-\varepsilon_1$. Then there is some $\delta > 0$ with

$$\mathbb{P}\Big[-\frac{1}{n_k}\log\pi_{n_k}(X_{n_k}) \geq \frac{1}{N}\Big] \geq \delta \quad \forall k \in \mathbb{N} \text{ large enough.}$$

To see this, assume that $\delta = \delta_k$ depends on k with $\liminf_{k \to \infty} \delta_k = 0$ which leads to a contradiction to (7.2) since

$$(-\log p_0) \cdot \delta_k + (1 - \delta_k) \frac{1}{N} \geq -\frac{1}{n_k} \mathbb{E}\big[\log \pi_{n_k}(X_{n_k})\big] \geq c - \varepsilon_1.$$

If δ_k tends to zero then we get a contradiction to the choice of N.

Choose now $\varepsilon > 0$ arbitrarily small with $\varepsilon < \delta$. In the recurrent case we have $\ell = 0$. Then there is some index $K \in \mathbb{N}$ such that for all $k \geq K$:

$$\delta - \varepsilon \leq \mathbb{P}\big[-\log \pi_{n_k}(X_{n_k}) \geq n_k/N, |X_n| \leq \varepsilon n_k\big] \leq e^{-n_k/N} \cdot |\mathcal{A}|^{\varepsilon n_k}$$

which yields the inequality

$$\frac{1}{N} + \frac{1}{n_k} \log(\delta - \varepsilon) \leq \varepsilon \log |\mathcal{A}|.$$

But this gives a contradiction if we make ε sufficiently small since the right hand side tends to zero, but the left hand side to $\frac{1}{N}$. Thus, $\limsup_{n \to \infty} -\frac{1}{n} \mathbb{E}\big[\log \pi_n(X_n)\big] = 0$, yielding $h = 0$. $\qquad\square$

8. Analyticity of Entropy

The random walk on $\mathcal{L}$ depends on *finitely* many parameters which are described by the transition probabilities $p(w_1, w_2)$, $w_1, w_2 \in \mathcal{A}^*$ with $|w_1| \leq 2$ and $|w_2| \leq 3$. That is, each random walk on $\mathcal{L}$ can be defined via a vector $\underline{p} \in \mathbb{R}^{|\mathcal{B}_2 \times \mathcal{B}_3|}$, where $\mathcal{B}_i := \cup_{n=1}^{i} \mathcal{A}^n \cup \{o\}$. The support of $\underline{p}$ is the set of indices in $\mathcal{B}_2 \times \mathcal{B}_3$ corresponding to non-zero entries of $\underline{p}$. Fix now any subset $B \subseteq \mathcal{B}_2 \times \mathcal{B}_3$, which allows at least one well-defined random walk on $\mathcal{A}^*$, and consider in the following only vectors $\underline{p}$ with support B, which give rise to a well-defined random walk on $\mathcal{A}^*$. We ask whether the entropy mapping $\underline{p} \mapsto h = h_{\underline{p}}$ varies real-analytically. The crucial point will be the following lemma:

Lemma 8.1. *The transition probabilities $q(w_1, w_2)$, $w_1, w_2 \in \mathcal{W}_0$, vary real-analytically w.r.t. $\underline{p}$.*

Proof. Observe that analyticity of $q(w_1, w_2)$ follows from analyticity of $\xi(ab)$, $H(ab, c)$, $ab \in \mathcal{A}^2$, $c \in \mathcal{A}$ and $\bar{L}(ab, cde)$, $de \in \mathcal{A}^2$. Hence, we prove real-analyticity of these generating functions. The function $z \mapsto H(ab, c|z)$ has radius of convergence bigger than 1, which can be easily deduced from Lemma 6.1. Thus, for $\delta > 0$ small enough, we have

$$\infty > H(ab, c|1 + \delta) = \sum_{n \geq 1} \mathbb{P}_{ab}[X_n = c, \forall m < n : |X_m| \geq 2](1 + \delta)^n.$$

The probability $\mathbb{P}_{ab}[X_n = c, \forall m < n : |X_m| \geq 2]$ can be rewritten as

$$\sum_{\substack{n_1, \ldots, n_d \geq 1: \\ n_1 + \cdots + n_d = n}} c(n_1, \ldots, n_d) p_1^{n_1} \cdot \ldots \cdot p_d^{n_d},$$

where $p_1, \ldots, p_d$ correspond to the non-zero entries of the vector $\underline{p}$. Therefore,

$$H(ab, c|1 + \delta) = \sum_{n \geq 1} \sum_{\substack{n_1, \ldots, n_d \geq 1: \\ n_1 + \cdots + n_d = n}} c(n_1, \ldots, n_d)(p_1(1 + \delta))^{n_1} \cdot \ldots \cdot (p_d(1 + \delta))^{n_d} < \infty.$$

Hence, $\underline{p}$ lies in the interior of the domain of convergence of $H(ab, c|1)$ if seen as a multivariate power series in terms of $\underline{p}$. This yields real-analyticity of $H(ab, c|1)$ in $\underline{p}$. The same holds for $\xi(ab)$ and $\bar{L}(ab, cde)$, which is proven completely analogously since $\bar{\bar{L}}(ab, cde|z)$ has also radius of convergence bigger than 1, see proof of Lemma 6.1. $\qquad\square$

Now we can prove:

Theorem 8.2. *The entropy h varies real-analytically under all probability measures of constant support.*

Proof. The claim follows now easily via the equation $h = \ell \cdot \lambda^{-1} \cdot H(\mathbf{Z})$. By Lemma 8.1, ν_0 (as the solution of a linear system of equations in terms of $q(\cdot, \cdot)$ is real-analytic, so λ is analytic. Moreover, by Han and Marcus [10, Theorem 1.1], $H(\mathbf{Z})$ is also real-analytic. Real-analyticity of ℓ can be shown completely analogously to the proof of Lemma 8.1 with the help of the formula for ℓ given in [6, Theorem 2.4]. This finishes the proof. $\qquad\square$

REFERENCES

[1] A. Avez. Entropie des groupes de type fini. *C.R.Acad.Sci.Paris Sér. A-B*, 275:A1363–A1366, 1972.
[2] I. Benjamini and Y. Peres. Tree-indexed random walks on groups and first passage percolation. *Probab. Theory Related Fields*, 98(1):91–112, 1994.
[3] S. Blachère, P. Haïssinsky, and P. Mathieu. Asymptotic entropy and Green speed for random walks on countable groups. *Ann. Probab.*, 36(3):1134–1152, 2008.
[4] T. Cover and J. Thomas. *Elements of Information Theory.* Wiley & Sons, 2nd edition, 2006.
[5] L. Gilch and F. Ledrappier. Regularity of the drift and entropy of random walks on groups. *Publ. Mat. Urug.*, to appear.
[6] L. A. Gilch. Rate of escape of random walks on regular languages and free products by amalgamation of finite groups. *Discrete Math. Theor. Comput. Sci. Proc.*, pages 405–420, 2008.
[7] L. A. Gilch. Asymptotic entropy of random walks on free products. *Electron. J. Probab.*, 16:76–105, 2011.
[8] L. A. Gilch and S. Müller. Random walks on directed covers of graphs. *J. Theoret. Probab.*, 24(1):118–149, 2009.
[9] P. Haïssinsky, P. Mathieu, and S. Müller. Renewal theory for random walks on cocompact Fuchsian groups. *Preprint*, 2012.
[10] G. Han and B. Marcus. Analyticity of entropy rate of hidden Markov chains. *IEEE Trans. Inform. Theory*, 52(12):5251–5266, 2006.
[11] J. Kingman. The ergodic theory of subadditive processes. *J. Roy. Statist. Soc., Ser. B*, 30:499–510, 1968.
[12] S. Lalley. Random walks on regular languages and algebraic systems of generating functions. *Algebraic Methods in Statistics and Probability, Contemp. Math.*, 287(201–230), 2000.
[13] F. Ledrappier. Analyticity of the entropy for some random walks. *Groups Geom. Dyn.*, 6:317–333, 2012.
[14] F. Ledrappier. Regularity of the entropy for random walks on hyperbolic groups. *Ann. Probab.*, to appear.
[15] V. Malyshev. Stabilization laws in the evolution of a random string. *Problems Inform. Transmission*, 30:260–274, 1995.

[16] T. Nagnibeda and W. Woess. Random walks on trees with finitely many cone types. *J. Theoret. Probab.*, 15:399–438, 2002.

Lorenz Gilch: Graz University of Technology, Steyrergasse 30/III/C304, 8010 Graz, Austria

E-mail address: gilch@TUGraz.at

URL: http://www.math.tugraz.at/∼gilch/