

# How long is the Chaos Game?

Ian D. Morris and Natalia Jurga

*Ian Morris thanks his former teacher, John Little, for introducing him to fractals in general and the chaos game in particular at New College, Swindon, in the 1995-96 academic year.*

## ABSTRACT

In the 1988 textbook *Fractals Everywhere* M. Barnsley introduced an algorithm for generating fractals through a random procedure which he called the *chaos game*. Using ideas from the classical theory of covering times of Markov chains we prove an asymptotic formula for the expected time taken by this procedure to generate a  $\delta$ -dense subset of a given self-similar fractal satisfying the open set condition.

## 1. Introduction

An *iterated function system* or *IFS* is defined to be a tuple of contracting transformations of a complete metric space, which in this article will be taken to be  $\mathbb{R}^d$ . It is well-known that if  $(S_1, \dots, S_N)$  is such an IFS then there exists a unique non-empty, compact set  $F = \bigcup_{i=1}^N S_i F \subseteq \mathbb{R}^d$  which is called the *attractor* or *limit set* of  $(S_1, \dots, S_N)$ . The properties of attractors of iterated function systems and the natural measures supported on them have been the subject of substantial mathematical inquiry for several decades since their introduction in [3, 8], and remain a highly active topic of contemporary mathematical research (we note for example [1, 4, 7, 10, 13, 14, 15]) as well as being noted for their aesthetic appeal.

In [2] Barnsley introduced an algorithm known as the *Chaos Game* for the construction of the limit set  $F$  of an iterated function system  $(S_1, \dots, S_N)$ . Given an arbitrary starting point  $x_0 \in \mathbb{R}^d$ , we define a sequence  $(x_n)_{n=0}^\infty$  inductively by choosing for each  $n \geq 1$  an index  $i_n \in \{1, \dots, N\}$  independently at random according to some fixed non-degenerate probability vector  $(p_1, \dots, p_N)$ , and taking  $x_n := S_{i_n} x_{n-1}$  for every  $n \geq 1$ . It is not difficult to show that the resulting sequence almost surely has the attractor  $F$  as its  $\omega$ -limit set (that is, we have  $\bigcap_{m=1}^\infty \overline{\{x_n : n \geq m\}} = F$ ) and it is not much more difficult to show that the distribution  $\frac{1}{n} \sum_{k=0}^{n-1} \delta_{x_k}$  converges almost surely to the unique Borel probability measure  $m$  supported on  $F$  which satisfies  $m = \sum_{i=1}^N p_i (S_i)_* m$ , as was first established in [5]. If the initial point  $x_0$  is taken to be in the attractor (for example, by taking  $x_0$  to be the fixed point of one of the contractions  $S_i$ ) then one obtains the simpler result that the sequence  $(x_n)_{n=0}^\infty$  is almost surely dense in the attractor, and for the rest of this article we will prefer to make this assumption on the starting point  $x_0$ . Yet surprisingly, we have found no trace in the literature of the following question: *how quickly does the randomly-generated sequence  $(x_n)$  become dense in the attractor?* In this direction we are aware only of the article [6], which informally investigates the problem of choosing probabilities in such a way as to generate fractal images with maximal efficiency using the chaos game procedure. In the present note we attempt to fill this gap in the literature with a rigorous investigation.

Let us make our question precise. Given a compact subset  $F$  of  $\mathbb{R}^d$  we will say that a subset  $X$  of  $F$  is  $\delta$ -dense in  $F$  if for every  $z \in F$  there exists  $x \in X$  such that  $d(x, z) \leq \delta$  in the standard metric on  $\mathbb{R}^d$ . (Since  $X$  is a subset of  $F$ , this is equivalent to asking that the Hausdorff distance between  $X$  and  $F$  is at most  $\delta$ .) Given an IFS  $(S_1, \dots, S_N)$  with attractor  $F$ , for each  $\delta > 0$ ,  $\mathbf{i} = i_1 i_2 \dots \in \{1, \dots, N\}^\mathbb{N}$  and starting point  $v \in F$  we define the  $\delta$ -waiting time along the sequence  $\mathbf{i}$  as

$$W_{\delta,v}(\mathbf{i}) := \inf \{n \geq 1 : \{S_{i_1} v, S_{i_2} S_{i_1} v, \dots, S_{i_n} \cdots S_{i_1} v\} \text{ is } \delta\text{-dense in } F\}.$$

If additionally a nondegenerate probability vector  $(p_1, \dots, p_N)$  is understood, then we define the *expected  $\delta$ -waiting time* with starting point  $v$  to be the expectation  $\mathbb{E}(W_{\delta,v})$  with respect to the  $(p_1, \dots, p_N)$ -Bernoulli measure on  $\{1, \dots, N\}^{\mathbb{N}}$ .

We recall that an IFS is said to satisfy the *open set condition* or *OSC* if there exists a nonempty open set  $U \subset \mathbb{R}^d$  such that  $\bigcup_{i=1}^N S_i U \subseteq U$  with the sets  $S_i U$  being pairwise disjoint. The set  $U$  may without loss of generality be taken to be bounded, and we will always assume that this is the case. We recall that  $S_i: \mathbb{R}^d \rightarrow \mathbb{R}^d$  is called a *similitude* or *similarity transformation* if there exists  $r_i \in (0, 1)$  such that  $d(S_i u, S_i v) = r_i d(u, v)$  for all  $u, v \in \mathbb{R}^d$ ; in this case we say that  $r_i$  is the *contraction ratio* of  $S_i$ . If  $(S_1, \dots, S_N)$  is an IFS of similitudes with respective contraction ratios  $r_1, \dots, r_N$  then the *similarity dimension* of  $(S_1, \dots, S_N)$  is defined to be the unique real number  $s \geq 0$  such that  $\sum_{i=1}^N r_i^s = 1$ . By a classical theorem of Hutchinson (see [8]), if an IFS of similitudes satisfies the open set condition then the Hausdorff and box dimensions of the attractor  $F$  are both equal to the similarity dimension  $s$ . It was also shown by Hutchinson that there exists a unique Borel probability measure  $m$  satisfying  $m = \sum_{i=1}^N r_i^s (S_i)_* m$ , and that this measure is supported on  $F$  and has Hausdorff dimension equal to that of  $F$ ; moreover, if any other probability vector is chosen then the resulting measure has Hausdorff dimension smaller than that of  $F$ . At an intuitive level this suggests that the limit distribution of the random sequence  $(x_n)_{n=0}^{\infty}$  generated by the Chaos Game will be most evenly distributed around the attractor when the underlying probability vector is  $(r_1^s, \dots, r_N^s)$ , and will be more concentrated in certain subregions of the attractor for other probability vectors. Thus we might expect the probability vector  $(r_1^s, \dots, r_N^s)$  to generate a random sequence which fills up the attractor most efficiently, and for other choices of probability vector to result in longer waiting times for the sequence to become  $\delta$ -dense in the attractor. This intuition is realised in our main result:

**THEOREM 1.1.** *Let  $(S_1, \dots, S_N)$  be an IFS of similitudes  $S_i: \mathbb{R}^d \rightarrow \mathbb{R}^d$ , with contraction ratios given by  $r_i \in (0, 1)$ , which satisfies the OSC. Let  $s$  denote the similarity dimension of  $(S_1, \dots, S_N)$ , let  $(p_1, \dots, p_N)$  be a nondegenerate probability vector, and define*

$$t := \max_{1 \leq i \leq N} \frac{\log p_i}{\log r_i}, \quad (1.1)$$

where we observe that by the arithmetic-geometric mean inequality we have  $t \geq s$  with equality if and only if  $(p_1, \dots, p_N) = (r_1^s, \dots, r_N^s)$ . If the maximum in (1.1) is attained at a unique value of  $i \in \{1, \dots, N\}$  then there exists a constant  $C > 0$  such that for every starting point  $v \in F$  and  $0 < \delta < \min_{1 \leq i \leq N} r_i$ ,

$$C^{-1} \delta^{-t} \log \log \left( \frac{1}{\delta} \right) \leq \mathbb{E}(W_{\delta,v}) \leq C \delta^{-t} \left( \log \log \left( \frac{1}{\delta} \right) \right)^2. \quad (1.2)$$

If the maximum in (1.1) is not attained at a unique value of  $i \in \{1, \dots, N\}$  then there exists a constant  $C > 0$  such that for every starting point  $v \in F$  and  $0 < \delta < \min_{1 \leq i \leq N} r_i$ ,

$$C^{-1} \delta^{-t} \log \left( \frac{1}{\delta} \right) \leq \mathbb{E}(W_{\delta,v}) \leq C \delta^{-t} \log \left( \frac{1}{\delta} \right). \quad (1.3)$$

Thus for the probability vector  $(p_1, \dots, p_N) = (r_1^s, \dots, r_N^s)$  we have for every starting point  $v \in F$

$$C^{-1} \delta^{-s} \log \left( \frac{1}{\delta} \right) \leq \mathbb{E}(W_{\delta,v}) \leq C \delta^{-s} \log \left( \frac{1}{\delta} \right)$$

for all  $0 < \delta < \min_{1 \leq i \leq N} r_i$ , and for every other probability vector  $\mathbb{E}(W_{\delta,v})$  tends to infinity more rapidly as  $\delta \rightarrow 0$ . At an intuitive level the principle underlying this result is that for the “natural” probability measure  $(r_1^s, \dots, r_N^s)$ , all regions of the attractor with diameter  $\delta$  take an approximately equal time to visit; for other probability measures, some  $\delta$ -balls in the attractor are substantially more difficult to access than others. As we will see below, the key determiner of the expected waiting time is the expected time taken to visit the most slowly accessible part of the attractor, and it transpires that this in turn corresponds to a region of the form  $S_i^n U$  where  $i$  is chosen to maximise the ratio  $\log p_i / \log r_i$  and  $n \geq 1$  is chosen such that this region has diameter approximately  $\delta$ . In the case where  $\log p_i / \log r_i$  is maximised at a unique index  $i$  it is interesting to ask to what extent the result (1.2) may be sharpened, but we have not been able to determine the exact rate of growth of  $\mathbb{E}(W_{\delta,v})$  in that case in the present article. It is

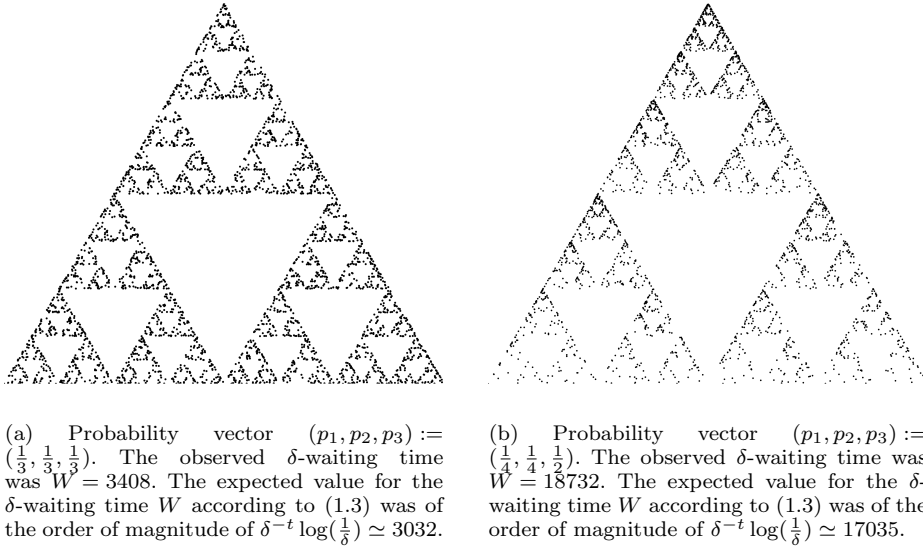


Figure 1: Each of these diagrams shows a randomly-generated sequence  $(x_n)_{n=0}^W$  starting at  $x_0 := 0$  and terminating when every point of the Sierpinski triangle has been approached to within distance  $\delta := 2^{-6}$ . Here the underlying iterated function system is given by  $T_1(x) := \frac{x}{2}$ ,  $T_2(x) := \frac{x}{2} + (\frac{1}{2}, 0)$  and  $T_3(x) := \frac{x}{2} + (\frac{1}{4}, \frac{\sqrt{3}}{4})$ . As indicated by Theorem 1.1 and suggested by the accompanying heuristic description, with unbalanced probabilities the trajectory spends much more time confined to a small region closest to the fixed point of the transformation carrying highest probability weight, and requires correspondingly more points in order to terminate.

also interesting to ask what information may be obtained regarding the pointwise almost sure behaviour of the family of random variables  $W_{\delta,v}$  for fixed  $v$ .

In the case where the starting point  $v$  is not taken to be in the attractor, since the sequence  $S_{i_n} \cdots S_{i_1} v$  approaches the attractor at a uniform exponential rate, one may obtain the same asymptotics for the expected waiting time as in Theorem 1.1 but with a larger constant  $C$  depending on the initial distance between  $v$  and the attractor; we leave the details of this adaptation of our result to the reader.

To illustrate the structure of the proof of Theorem 1.1 it is helpful to consider a simpler case in which the contraction ratios  $r_i$  are all equal to the same constant  $r$  and the diameter of the set  $U$  is precisely 1. In this case, if every set of the form  $S_{i_m} \cdots S_{i_1} U$  has been visited by the sequence  $(x_n)_{n=0}^\infty$  by time  $N$  then we certainly have  $W_{r^m \kappa, v} \leq N$ . On the other hand one may show that there exists  $\kappa > 0$  such that for every  $m \geq 1$ , every set of the form  $S_{i_m} \cdots S_{i_1} U$  contains an open ball of radius  $r^m \kappa$  which in particular does not intersect any *other* set of the form  $S_{i_m} \cdots S_{i_1} U$ . Thus if the sequence of indices  $\mathbf{i}$  fails to include a particular string of the form  $i_1 \cdots i_m$  before time  $N$ , we expect that  $W_{r^m \kappa, v} > N$ . (There is some imprecision here in that the initial point  $v$  may by chance have belonged to the ball  $r^m \kappa$ , but it transpires that this imprecision has a negligible effect in practice.) This suggests that the asymptotic behaviour of the expectation of  $W_{\delta,v}$  can be reduced to the problem of determining the expected first time for an IID random sequence in  $\{1, \dots, N\}^N$ , chosen with respect to the Bernoulli measure  $(p_1, \dots, p_N)^N$ , to include all of the distinct words of length  $m$  over the alphabet  $\{1, \dots, N\}$ , where  $m$  is chosen so that  $r^m$  is approximately the size of  $\delta$ . But this symbolic problem is precisely the classical *coupon collector's problem* described in, for example, [9]. In the full generality of Theorem 1.1 this

approach must be adapted somewhat: since the sets  $S_{i_m} \cdots S_{i_1} U$  will in general have different diameters for different sequences  $i_1, \dots, i_m$  of the same length  $m$ , it is necessary to partition the set  $\{1, \dots, N\}^{\mathbb{N}}$  into cylinders and estimate the expected time for all of these cylinders to be visited by a random sequence. This results in a Markov chain analogue of the coupon collector's problem which we solve using techniques adapted from [9, §11].

Our proof of Theorem 1.1 will thus be divided into two parts: the reduction of the problem to a covering problem for Markov chains, and the solution of the latter covering problem. Some possible directions for future research are described at the end of this note.

## 2. A Markov chain construction

For the remainder of this article we fix an IFS of similarities  $(S_1, \dots, S_N)$  which satisfies the OSC and denote the contraction ratio of each map  $S_i$  by  $r_i$ . We also fix a nondegenerate probability vector  $(p_1, \dots, p_N)$ . Let us write  $r_{\min} := \min_{i=1, \dots, N} r_i$ . In this section we will show that for each  $0 < \delta < r_{\min}$ , we can construct a Markov chain whose expected time to visit all of its states is approximately proportional to  $\mathbb{E}W_{\delta, v_0}$ . More precisely, given a Markov chain  $(X_n)_{n=0}^{\infty}$  on a finite state space  $\Omega$ , we define the *covering time* by

$$\tau_{\text{cov}} = \min\{t \geq 0 : \forall y \in \Omega, \exists s \leq t \text{ s.t. } X_s = y\},$$

that is, the first time that all of the states in  $\Omega$  have been visited by the Markov chain. Given  $x \in \Omega$ , we denote by  $\mathbb{E}_x \tau_{\text{cov}}$  the expected covering time given that  $X_0 = x$ . We can now state the main result of this section.

**PROPOSITION 2.1.** *For each  $0 < \delta < r_{\min}$  there exists an irreducible Markov chain  $(X_n)_{n=0}^{\infty}$  on a finite state space  $P_{\delta}$  such that for each  $v_0 \in F$ , there exists  $\mathbf{i}_0 \in P_{\delta}$  for which*

$$\mathbb{E}W_{\frac{\delta}{c}, v_0} \leq \mathbb{E}_{\mathbf{i}_0} \tau_{\text{cov}} \leq \mathbb{E}W_{c\delta, v_0} \quad (2.1)$$

where the constant  $c \in (0, 1)$  is independent of  $\delta$ ,  $v_0$  and  $\mathbf{i}_0$ .

We will derive Proposition 2.1 from the combination of two results to be proved below, Proposition 2.2 and Proposition 2.3. The significance of (2.1) is that for each  $0 < \delta < r_{\min}$  and  $\mathbf{i}_0 \in P_{\delta}$ ,  $\mathbb{E}_{\mathbf{i}_0} \tau_{\text{cov}}$  can be estimated by employing classical methods for bounding covering times of irreducible Markov chains. In particular, we will show that for any  $0 < \delta < r_{\min}$  and  $\mathbf{i}_0 \in P_{\delta}$ ,  $\mathbb{E}_{\mathbf{i}_0} \tau_{\text{cov}}$  satisfies the upper and lower bounds presented in (1.2) and (1.3), which will directly imply the bounds for  $\mathbb{E}W_{\delta, v_0}$  by (2.1), up to a change in the uniform constant.

We introduce some notation. Define  $\mathcal{I} = \{1, \dots, N\}$ , which we call the *index set*. Let  $\mathcal{I}^n = \{i_1 \dots i_n : i_j \in \mathcal{I}\}$  denote the set of all words of length  $n$  over the index set,  $\mathcal{I}^* = \bigcup_{n \in \mathbb{N}} \mathcal{I}^n$  the set of all finite words over the index set and  $\Sigma = \mathcal{I}^{\mathbb{N}}$  the set of all sequences over the index set.  $\Sigma$  is equipped with the infinite product topology with respect to which it is compact and metrisable. If  $\mathbf{i} \in \mathcal{I}^*$  and  $\mathbf{j} \in \mathcal{I}^* \cup \Sigma$  we let  $\mathbf{i}\mathbf{j}$  denote the concatenation of  $\mathbf{i}$  with  $\mathbf{j}$ . Given  $\mathbf{i} \in \mathcal{I}^*$ , let  $[\mathbf{i}]$  denote the cylinder set  $[\mathbf{i}] = \{\mathbf{i}\mathbf{j} : \mathbf{j} \in \Sigma\}$ . Cylinder sets are clopen and generate the topology on  $\Sigma$ . Given  $\mathbf{i} = i_1 \dots i_n \in \mathcal{I}^*$  let  $|\mathbf{i}|$  denote the length of the word  $\mathbf{i}$ , so that in this case  $|\mathbf{i}| := n$ . Given  $\mathbf{i} = i_1 \dots i_n \in \mathcal{I}^*$  with  $|\mathbf{i}| \geq 2$  denote  $\mathbf{i}^- := i_1 \dots i_{n-1}$ . Given  $\mathbf{i} = i_1 i_2 \dots \in \Sigma$  or  $\mathbf{i} = i_1 \dots i_{n+m} \in \mathcal{I}^*$  let  $|\mathbf{i}|_n := i_1 \dots i_n$ . Given  $\mathbf{i} = i_1 \dots i_n \in \mathcal{I}^*$  let  $r_{\mathbf{i}} := r_{i_1} \cdots r_{i_n}$ ,  $p_{\mathbf{i}} := p_{i_1} \cdots p_{i_n}$  and  $S_{\mathbf{i}} := S_{i_1} \cdots S_{i_n}$ .

It is not difficult to show that for every  $\mathbf{i} \in \Sigma$  the limit

$$\pi(\mathbf{i}) := \lim_{n \rightarrow \infty} S_{i_1} \cdots S_{i_n} v$$

exists for every  $v \in \mathbb{R}^d$  and that moreover the limit is independent of the starting point  $v$ . This *coding map*  $\pi : \Sigma \rightarrow \mathbb{R}^d$  is continuous and its image  $\pi(\Sigma)$  is precisely  $F$ . (However, in cases where  $\overline{T_i U} \cap \overline{T_j U} \neq \emptyset$  for some  $i \neq j$  the coding map  $\pi$  can fail to be injective.) Obviously, for every  $\mathbf{i} \in \Sigma$  and  $n \geq 1$  we have  $\mathbf{i} \in [\mathbf{j}]$  where  $\mathbf{j}$  is the word corresponding to the first  $n$  symbols of  $\mathbf{i}$ . Since  $\pi : \Sigma \rightarrow F$  is surjective, for every  $x \in F$  and  $n \geq 1$  there exists at least one word  $\mathbf{j} \in \mathcal{I}^*$  with  $|\mathbf{j}| = n$  such that  $x \in \pi([\mathbf{j}]) = S_{\mathbf{j}} F$ .

For each  $\delta \in (0, r_{\min})$ , define a subset of  $\mathcal{I}^*$  by

$$P_{\delta} = \{\mathbf{i} \in \mathcal{I}^* : r_{\mathbf{i}} \leq \delta < r_{\mathbf{i}^-}\}.$$

Note that since  $\delta < r_{\min}$ , if  $r_i \leq \delta$  then necessarily  $|\mathbf{i}| \geq 2$  and hence  $\mathbf{i}^-$  is well defined. It is easy to see that  $\{[\mathbf{i}] : \mathbf{i} \in P_\delta\}$  is a finite partition of  $\Sigma$ . Therefore, for any  $x \in F$  there exists at least one word  $\mathbf{i} \in P_\delta$  such that  $x \in S_{\mathbf{i}}F$ . Let  $N_\delta = |P_\delta|$ , where  $|P_\delta|$  denotes the cardinality of the set  $P_\delta$ . We claim that  $\delta^{-s} \leq N_\delta \leq r_{\min}^{-s} \delta^{-s}$ . To see this let  $\mathbb{P}$  denote the Bernoulli probability measure on  $\Sigma$  defined by  $\mathbb{P}([\mathbf{i}]) = r_{\mathbf{i}}^s$  for every  $\mathbf{i} \in \mathcal{I}^*$ . Then  $\sum_{\mathbf{i} \in P_\delta} \mathbb{P}([\mathbf{i}]) = 1$  since  $\{[\mathbf{i}] : \mathbf{i} \in P_\delta\}$  is a partition, which is to say  $\sum_{\mathbf{i} \in P_\delta} r_{\mathbf{i}}^s = 1$ . Therefore  $r_{\min}^s \delta^s N_\delta < 1 \leq \delta^s N_\delta$  which proves the claim.

We say that a list of words  $\mathbf{j}_1, \dots, \mathbf{j}_k \in \mathcal{I}^*$  visits the cylinder set  $[\mathbf{i}] \subset \Sigma$  if at least one of the words  $\mathbf{j}_i$  satisfies  $[\mathbf{j}_i] \subseteq [\mathbf{i}]$ . We will show that instead of keeping track of which regions of the attractor are visited by the chaos game algorithm, we can keep track of which cylinder sets in  $\{[\mathbf{i}] : \mathbf{i} \in P_\delta\}$  are visited by a symbolic analogue of the algorithm. This is made precise in the following proposition.

**PROPOSITION 2.2.** *There exist  $\kappa > 0$  and  $\Delta > 0$  depending only on  $(S_1, \dots, S_N)$  having the following property. Let  $v_0 \in F$  be arbitrary,  $\delta \in (0, r_{\min})$  and choose any  $\mathbf{i}_0 \in P_\delta$  such that  $v_0 \in S_{\mathbf{i}_0}F$ :*

- (i) *If  $\mathbf{i}_0, \mathbf{i}_1 \mathbf{i}_0, \dots, \mathbf{i}_n \dots \mathbf{i}_1 \mathbf{i}_0$  visits every cylinder in  $\{[\mathbf{i}] : \mathbf{i} \in P_\delta\}$  then the set  $\{v_0, S_{\mathbf{i}_1} v_0, \dots, S_{\mathbf{i}_n} \dots S_{\mathbf{i}_1} v_0\}$  is  $2\delta\Delta$ -dense in  $F$ .*
- (ii) *If  $\{v_0, S_{\mathbf{i}_1} v_0, \dots, S_{\mathbf{i}_n} \dots S_{\mathbf{i}_1} v_0\}$  is  $\kappa\delta$ -dense in  $F$  then  $\mathbf{i}_0, \mathbf{i}_1 \mathbf{i}_0, \dots, \mathbf{i}_n \dots \mathbf{i}_1 \mathbf{i}_0$  visits each cylinder in  $\{[\mathbf{i}] : \mathbf{i} \in P_\delta\}$ .*

*Proof.* Since the IFS  $(S_1, \dots, S_N)$  satisfies the OSC, by a result of A. Schief ([12]) it also satisfies the strong open set condition (SOSC), that is, there exists a bounded open set  $U$  such that  $\bigcup_{i=1}^N S_i U \subset U$  where the union is disjoint and  $U \cap F \neq \emptyset$ . Let  $\Delta := \text{diam}F$ . It follows that there exists  $x \in F$  and  $0 < \epsilon < \Delta$  with  $B(x, \epsilon) \subset U$ . In particular for  $\mathbf{i}, \mathbf{j} \in P_\delta$  with  $\mathbf{i} \neq \mathbf{j}$ , the balls  $S_{\mathbf{i}}B(x, \epsilon)$  and  $S_{\mathbf{j}}B(x, \epsilon)$  are disjoint. If  $y \in F$  is arbitrary then we may choose  $\mathbf{i} \in P_\delta$  such that  $y \in S_{\mathbf{i}}F$ . Since  $S_{\mathbf{i}}x, y \in S_{\mathbf{i}}F$  it then follows that  $|S_{\mathbf{i}}x - y| \leq r_{\mathbf{i}}\Delta$ . We have shown that

$$F \subset \bigcup_{\mathbf{i} \in P_\delta} B(S_{\mathbf{i}}x, r_{\mathbf{i}}\Delta). \quad (2.2)$$

Now, to prove (i), fix  $\mathbf{i}_0, \mathbf{i}_1 \mathbf{i}_0, \dots, \mathbf{i}_n \dots \mathbf{i}_1 \mathbf{i}_0$  satisfying the hypothesis of (i) and consider an arbitrary  $\mathbf{i} \in P_\delta$ . By assumption there exists  $k$  satisfying  $0 \leq k \leq n$  such that  $[\mathbf{i}_k \dots \mathbf{i}_1 \mathbf{i}_0] \subset [\mathbf{i}]$ . Thus  $S_{\mathbf{i}_k} \dots S_{\mathbf{i}_1} v_0 \in S_{\mathbf{i}}F$ , and in particular  $S_{\mathbf{i}_k} \dots S_{\mathbf{i}_1} v_0 \in B(S_{\mathbf{i}}x, r_{\mathbf{i}}\Delta)$ . Moreover, since for any  $y \in B(S_{\mathbf{i}}x, r_{\mathbf{i}}\Delta)$  we have

$$|y - S_{\mathbf{i}_k} \dots S_{\mathbf{i}_1} v_0| \leq |y - S_{\mathbf{i}}x| + |S_{\mathbf{i}}x - S_{\mathbf{i}_k} \dots S_{\mathbf{i}_1} v_0| < r_{\mathbf{i}}\Delta + r_{\mathbf{i}}\Delta \leq 2\delta\Delta,$$

it follows that  $B(S_{\mathbf{i}}x, r_{\mathbf{i}}\Delta) \subset B(S_{\mathbf{i}_k} \dots S_{\mathbf{i}_1} v_0, 2\delta\Delta)$ . Since  $\mathbf{i} \in P_\delta$  was arbitrary, it follows that

$$F \subset \bigcup_{\mathbf{i} \in P_\delta} B(S_{\mathbf{i}}x, r_{\mathbf{i}}\Delta) \subset \bigcup_{k=0}^n B(S_{\mathbf{i}_k} \dots S_{\mathbf{i}_1} v_0, 2\delta\Delta),$$

in other words,  $v_0, S_{\mathbf{i}_1} v_0, \dots, S_{\mathbf{i}_n} \dots S_{\mathbf{i}_1} v_0$  is  $2\delta\Delta$ -dense in  $F$  as required to establish (i).

To prove (ii), define  $\kappa := r_{\min}\epsilon$ . For each  $\mathbf{i} \in P_\delta$  we have  $S_{\mathbf{i}}x \in S_{\mathbf{i}}F \subset F$  and  $S_{\mathbf{i}}(B(x, \epsilon)) = B(S_{\mathbf{i}}x, r_{\mathbf{i}}\epsilon)$ , and we note also that  $\delta\kappa = \delta r_{\min}\epsilon \leq r_{\mathbf{i}}\epsilon$  where we have used the definition of  $P_\delta$ . Thus if  $\{v_0, S_{\mathbf{i}_1} v_0, \dots, S_{\mathbf{i}_n} \dots S_{\mathbf{i}_1} v_0\}$  is  $\kappa\delta$ -dense in  $F$  then for each  $\mathbf{i} \in P_\delta \setminus \{\mathbf{i}_0\}$  there must exist  $k$  satisfying  $0 \leq k \leq n$  such that  $S_{\mathbf{i}_k} \dots S_{\mathbf{i}_1} v_0$  is  $\kappa\delta$ -close to  $S_{\mathbf{i}}x$  and therefore satisfies  $S_{\mathbf{i}_k} \dots S_{\mathbf{i}_1} v_0 \in B(S_{\mathbf{i}}x, \kappa\delta) \subseteq B(S_{\mathbf{i}}x, r_{\mathbf{i}}\epsilon) = S_{\mathbf{i}}B(x, \epsilon)$ . By the definition of  $\epsilon$  we know that  $\{S_{\mathbf{i}}B(x, \epsilon)\}_{\mathbf{i} \in P_\delta}$  are disjoint balls, so we necessarily have  $[\mathbf{i}_k \dots \mathbf{i}_1 \mathbf{i}_0] \subset [\mathbf{i}]$  and therefore  $\mathbf{i}_0, \mathbf{i}_1 \mathbf{i}_0, \dots, \mathbf{i}_n \dots \mathbf{i}_1 \mathbf{i}_0$  visits every cylinder  $[\mathbf{i}]$  such that  $\mathbf{i} \in P_\delta$ . This completes the proof.  $\square$

Proposition 2.2 is key to the construction of the Markov chain  $(X_n^\delta)_{n=0}^\infty$ , which we are now ready to provide details of:

**PROPOSITION 2.3.** *Let  $\delta \in (0, r_{\min})$ . Define a square matrix  $A_\delta = [a_{\mathbf{i}, \mathbf{j}}]_{\mathbf{i}, \mathbf{j} \in P_\delta}$  of dimension  $N_\delta$  by*

$$a_{\mathbf{i}, \mathbf{j}} := \begin{cases} p_{\mathbf{i}} & \text{if } [\mathbf{i}\mathbf{j}] \subset [\mathbf{j}] \\ 0 & \text{otherwise,} \end{cases}$$

and define a vector  $\pi_\delta \in \mathbb{R}^{N_\delta}$  by  $\pi_{\mathbf{i}} := p_{\mathbf{i}}$ . Then:

- (a)  $A_\delta$  is a row stochastic matrix,
- (b)  $\pi_\delta$  is a left stationary vector, i.e.  $\pi_\delta A_\delta = \pi_\delta$ ,
- (c)  $A_\delta$  is irreducible.

In order to prove Proposition 2.3 we require two preliminary lemmas concerning  $P_\delta$ .

LEMMA 2.4. *Let  $\delta \in (0, r_{\min})$ . The Markov partition property holds: for every  $\mathbf{i} = i_1 \cdots i_n \in P_\delta$  there exist  $j_1, \dots, j_m \in P_\delta$  such that  $[\mathbf{i}] = \bigcup_{k=1}^m [i_1 j_k]$ .*

*Proof.* If  $n = 1$  then we have  $[\mathbf{i}] = [i_1] = \bigcup_{j \in P_\delta} [i_1 j]$  using the fact that  $\bigcup_{j \in P_\delta} [j] = \Sigma$ , so assume  $n \geq 2$ . Given  $\mathbf{i} = i_1 \cdots i_n \in P_\delta$  with  $n \geq 2$ , define  $\mathbf{i}' = i_2 \cdots i_n$ . Since  $\{[j] : j \in P_\delta\}$  is a partition there exist  $j_1, \dots, j_m \in P_\delta$  such that  $[\mathbf{i}'] \subseteq \bigcup_{k=1}^m [j_k]$  and  $[\mathbf{i}'] \cap [j_k] \neq \emptyset$  for all  $k = 1, \dots, m$ . We appeal to the fact that if two cylinder sets intersect then one of them contains the other. If for some  $k$  the set  $[\mathbf{i}']$  is a subset of  $[j_k]$  then either  $\mathbf{i}' = j_k$  or  $\mathbf{i}' = j_k \mathbf{k}$  for some finite word  $\mathbf{k}$ . In the former case we have  $[\mathbf{i}] = [i_1 \mathbf{i}'] = [i_1 j_k]$  and the proof is complete. In the latter case we have  $\mathbf{i} = i_1 \mathbf{i}' = i_1 j_k \mathbf{k}$  so that  $i_1 j_k$  is a proper prefix of  $\mathbf{i} \in P_\delta$ . This implies  $r_{i_1} r_{j_k} > \delta$  and in particular  $r_{j_k} > r_{i_1}^{-1} \delta > \delta$  contradicting that  $j_k \in P_\delta$ . We conclude that  $[j_k] \subseteq [i_1 \mathbf{i}']$  for each  $k = 1, \dots, m$ , so  $[\mathbf{i}'] \subseteq \bigcup_{k=1}^m [j_k] \subseteq [i_1 \mathbf{i}']$  and the result  $[\mathbf{i}] = \bigcup_{k=1}^m [i_1 j_k]$  follows.  $\square$

LEMMA 2.5. *Let  $\delta \in (0, r_{\min})$ . Then for every  $\mathbf{i} \in P_\delta$  and  $i \in \{1, \dots, N\}$  there exists a unique  $\mathbf{j} \in P_\delta$  such that  $[i\mathbf{i}] \subseteq [\mathbf{j}]$ .*

*Proof.* Fix such an  $i$  and  $\mathbf{i}$ . To demonstrate the existence of  $\mathbf{j}$  we observe that by the partition property there must exist  $\mathbf{j} \in P_\delta$  such that  $[i\mathbf{i}] \cap [\mathbf{j}] \neq \emptyset$ . Writing  $[\mathbf{j}] = \bigcup_{k=1}^m [i j_k]$  using the Markov partition property (Lemma 2.4) we see that there exists  $k$  such that  $[i\mathbf{i}] \cap [i j_k] \neq \emptyset$ . By the partition property this is only possible if  $\mathbf{i} = j_k$  and we deduce that  $[i\mathbf{i}] \subseteq [\mathbf{j}]$ . This proves existence. To obtain uniqueness we observe that if distinct  $j_1, j_2 \in P_\delta$  satisfy  $[i\mathbf{i}] \subseteq [j_1]$  and  $[i\mathbf{i}] \subseteq [j_2]$  then  $[j_1] \cap [j_2] \neq \emptyset$  and the partition property is contradicted. The lemma is proved.  $\square$

*Proof of Proposition 2.3.* To see that  $A_\delta$  is row stochastic we observe that for each  $\mathbf{i} \in P_\delta$  and each  $i = 1, \dots, N$  there exists a unique  $\mathbf{j} \in P_\delta$  such that  $[i\mathbf{i}] \subseteq [\mathbf{j}]$  by Lemma 2.5. This implies there exists a unique  $\mathbf{j} \in P_\delta$  such that  $a_{i,\mathbf{j}} = p_i$ . It follows that every  $p_1, \dots, p_N$  occurs once in the row of  $A_\delta$  corresponding to  $\mathbf{i}$  and the remaining entries in that column are zero, so  $A_\delta$  is row stochastic as claimed.

That  $\pi_\delta$  is a stochastic vector follows directly from the fact that  $P_\delta$  is a partition of  $\Sigma$ . To verify the equation  $\pi_\delta A_\delta = \pi_\delta$ , let  $\mathbf{j} \in P_\delta$  be arbitrary and using Lemma 2.4 we can write  $[\mathbf{j}] = \bigcup_{k=1}^m [i\mathbf{i}_k]$  where  $\mathbf{i}_1, \dots, \mathbf{i}_m \in P_\delta$  and  $i$  is the first symbol of  $\mathbf{j}$ . We obtain

$$\sum_{\mathbf{i} \in P_\delta} \pi_{\mathbf{i}} a_{\mathbf{i}, \mathbf{j}} = \sum_{\substack{\mathbf{i} \in P_\delta \\ [i\mathbf{i}] \subseteq [\mathbf{j}]}} p_{\mathbf{i}} p_{\mathbf{i}} = \sum_{k=1}^m p_{\mathbf{i}_k} p_{\mathbf{i}_k} = p_{\mathbf{j}} = \pi_{\mathbf{j}}$$

as required, where the penultimate equation follows from  $[\mathbf{j}] = \bigcup_{k=1}^m [i\mathbf{i}_k]$  and where we have used the fact that if  $[i\mathbf{i}] \subseteq [\mathbf{j}]$  then  $\mathbf{i}$  is necessarily equal to some  $\mathbf{i}_k$  by the partition property.

To show that  $A_\delta$  is irreducible, it is sufficient to show that there exists  $L > 0$  such that  $A_\delta^L$  is a positive matrix. We will show that this is true for  $L = L_\delta := \max_{\mathbf{i} \in P_\delta} \{|\mathbf{i}|\}$ . Let  $\mathbf{i}, \mathbf{j} \in P_\delta$  be arbitrary with  $\mathbf{j} = j_1, \dots, j_n$ , say, and if  $n < L$  let  $j_{n+1}, \dots, j_L \in \{1, \dots, N\}$  be arbitrary. Define  $\mathbf{k}_{L+1} := \mathbf{i}$ . By a simple inductive application of Lemma 2.5 starting at  $t = L$  and descending to  $t = 1$  we may choose  $\mathbf{k}_1, \dots, \mathbf{k}_L$  such that for all  $t = 1, \dots, L$  we have  $[j_t \mathbf{k}_{t+1}] \subseteq [\mathbf{k}_t]$ . Define  $\ell_t := \min\{L + 1 - t, |\mathbf{k}_t|\}$  for each  $t = 1, \dots, L + 1$ . We claim that for each  $t = 1, \dots, L + 1$  the first  $\ell_t$  symbols of  $\mathbf{k}_t$  are  $j_t, \dots, j_{t-1+\ell_t}$  in that order. This statement is clearly true for  $t = L + 1$ , so let us assume its truth for some  $t \in \{2, \dots, L + 1\}$  and deduce its truth for  $t - 1$ . Since  $[j_{t-1} \mathbf{k}_t] \subseteq [\mathbf{k}_{t-1}]$  we have  $|\mathbf{k}_{t-1}| \leq 1 + |\mathbf{k}_t|$  and therefore  $\ell_{t-1} = \min\{L + 2 - t, |\mathbf{k}_{t-1}|\} \leq \min\{L + 2 - t, 1 + |\mathbf{k}_t|\} = 1 + \ell_t$ . The relation  $[j_{t-1} \mathbf{k}_t] \subseteq [\mathbf{k}_{t-1}]$  also implies that the first  $|\mathbf{k}_{t-1}|$

symbols of  $j_{t-1}\mathbf{k}_t$  are precisely the word  $\mathbf{k}_{t-1}$ . Since the first  $1 + \ell_t$  symbols of  $j_{t-1}\mathbf{k}_t$  are  $j_{t-1}j_t \cdots j_{t-1+\ell_t}$ , the first  $\ell_{t-1} \leq \min\{|\mathbf{k}_{t-1}|, 1 + \ell_t\}$  symbols of  $\mathbf{k}_{t-1}$  must be  $j_{t-1} \cdots j_{t-2+\ell_{t-1}}$ . This is precisely what is required for the claim to be true in the case  $t - 1$ . The claim follows by induction.

Applying the claim with  $t = 1$  it follows that the first  $\ell_1 = \min\{L, |\mathbf{k}_1|\} = |\mathbf{k}_1|$  symbols of  $\mathbf{k}_1$  are  $j_1, \dots, j_{\ell_1}$ . If  $\ell_1 < n$  then  $[j]$  is a proper subset of  $[\mathbf{k}_1]$  and if  $\ell_1 > n$  then  $[\mathbf{k}_1]$  is a proper subset of  $[j]$ , but both of these contradict the partition property of  $P_\delta$  and we conclude that  $\ell_1 = n$  and therefore  $\mathbf{k}_1 = j$ . The relation  $[j_t\mathbf{k}_{t+1}] \subseteq [\mathbf{k}_t]$  for each  $t$  implies that  $a_{\mathbf{k}_{t+1}, \mathbf{k}_t} > 0$  for all  $t = 1, \dots, L$ . Since  $\mathbf{k}_1 = j$  and  $\mathbf{k}_{L+1} = \mathbf{i}$  it follows that  $a_{i, \mathbf{k}_L} a_{\mathbf{k}_L, \mathbf{k}_{L-1}} \cdots a_{\mathbf{k}_2, j} > 0$ . This is precisely what is needed to show that the entry of  $A_\delta^L$  in position  $(\mathbf{i}, j)$  is positive. Since  $\mathbf{i}$  and  $j$  were arbitrary the proof of (c) is complete.  $\square$

We may now prove Proposition 2.1:

*Proof of Proposition 2.1.* Let  $\kappa > 0$  be as given by Proposition 2.2 and for each  $\delta \in (0, r_{\min})$  let  $(X_n^\delta)_{n=0}^\infty$  be the irreducible Markov chain on the state space  $P_\delta$  which is induced by the transition matrix  $A_\delta$  defined in Proposition 2.3. Given  $v_0 \in F$ , choose any  $\mathbf{i}_0 \in P_\delta$  such that  $v_0 \in S_{\mathbf{i}_0}F$ . For every  $\mathbf{i} = i_1 i_2 \dots \in \Sigma$  define

$$\mathscr{W}_{\delta, \mathbf{i}_0}(\mathbf{i}) := \inf\{m \geq 1 : \forall j \in P_\delta \exists n \in \mathbb{N} \text{ such that } 1 \leq n \leq m \text{ and } [i_n \dots i_1 \mathbf{i}_0] \subset [j]\},$$

so that  $\mathscr{W}_{\delta, \mathbf{i}_0}(\mathbf{i}) \in \mathbb{N} \cup \{+\infty\}$ . If  $\mathscr{W}_{\delta, \mathbf{i}_0}(\mathbf{i}) = n \in \mathbb{N}$  then by definition  $\mathbf{i}_0, i_1 \mathbf{i}_0, \dots, i_n \cdots i_1 \mathbf{i}_0$  have visited all cylinders in  $\{[j] : j \in P_\delta\}$ . By Proposition 2.2(i) it follows that the set  $\{v_0, S_{i_1}v_0, \dots, S_{i_n} \cdots S_{i_1}v_0\}$  is  $2\delta\Delta$ -dense in  $F$ , so we have  $W_{2\delta\Delta, v_0}(\mathbf{i}) \leq n = \mathscr{W}_{\delta, \mathbf{i}_0}(\mathbf{i})$ . Similarly, we obtain  $\mathscr{W}_{\delta, \mathbf{i}_0}(\mathbf{i}) \leq W_{\kappa\delta, v_0}(\mathbf{i})$  from Proposition 2.2(ii). Hence

$$W_{2\delta\Delta, v_0}(\mathbf{i}) \leq \mathscr{W}_{\delta, \mathbf{i}_0}(\mathbf{i}) \leq W_{\kappa\delta, v_0}(\mathbf{i}). \quad (2.3)$$

By the definition of  $(X_n^\delta)_{n=0}^\infty$  it is clear that  $\mathbb{E}_{\mathbf{i}_0} \tau_{\text{cov}} = \mathbb{E} \mathscr{W}_{\delta, \mathbf{i}_0}$ , therefore it follows that Proposition 2.1 holds for  $(X_n^\delta)_{n=0}^\infty$  where the uniform constant  $c \in (0, 1)$  can be taken to be  $c = \min\{\frac{1}{2\Delta}, \kappa\}$ .  $\square$

### 3. Bounds on the covering time

Fix  $0 < \delta < r_{\min}$ . In order to prove Theorem 1.1 it suffices, via Proposition 2.1, to obtain estimates on the expected covering time  $\mathbb{E}_{\mathbf{i}_0} \tau_{\text{cov}}$  for all  $\mathbf{i}_0 \in P_\delta$ . To this end, we will make extensive use of a family of bounds on the expected covering time of irreducible Markov chains derived from the so called ‘‘Matthews method’’ [9, 11]. Roughly speaking, this type of bound reduces the task of estimating the expected covering time to only having to estimate the expected time for the Markov chain to travel between a pair of states. In particular, for  $\mathbf{i} \in P_\delta$  we define the hitting time

$$\tau_{\mathbf{i}} := \min\{t \geq 0 : X_t^\delta = \mathbf{i}\},$$

that is, the first time that the state  $\mathbf{i}$  is visited by the Markov chain, and denote by  $\mathbb{E}_{\mathbf{i}} \tau_j$  the expected hitting time of  $j \in P_\delta$  given that  $X_0^\delta = \mathbf{i}$ . Then, provided the expected hitting time between a pair of states is fairly homogeneous over a (large part of) the state space, the Matthews method exploits the fact that states can be visited by the Markov chain in many different orders to deduce that the expected covering time is approximately proportional to the logarithm of the cardinality of the state space times the typical expected hitting time between a pair of states. Translated through the symbolic coding introduced in the previous section, the requirement that the expected hitting time be fairly homogenous over a sufficiently large part of the state space turns out to correspond to the situation that the probability vector  $(p_1, \dots, p_N)$  is chosen in such a way that (1.1) is not satisfied for a unique index. In this situation the measures  $m = \sum_{i=1}^N p_i(S_i)_* m$  of many of the similarly sized pieces  $\{S_i F\}_{\mathbf{i} \in P_\delta}$  are approximately comparable, which implies that expected hitting times between many pairs of states in  $P_\delta$  are also roughly uniform. On the other hand, when the expected hitting time between pairs of states varies substantially according to the pair of states which are chosen – which is the case when (1.1) is satisfied uniquely and therefore the measure  $m$  is far from being uniformly distributed over any large subcollection of pieces  $\{S_i F\}_{\mathbf{i} \in P_\delta}$  – Matthews’ method yields bounds which are less

sharp, which accounts for the gap between the upper and lower estimates for the rate of growth of  $\mathbb{E}W_{\delta, v_0}$  in (1.3). We will provide precise statements for the ‘‘Matthews method’’ bounds which we use in Propositions 3.1, 3.3 and 3.7, each of which will appear directly preceding the proof of the corresponding bound from the part of Theorem 1.1 to which it pertains.

There are a couple more notions related to the covering and hitting times which will be useful in our analysis. Firstly, for  $\mathbf{i} \in P_\delta$  we define

$$\tau_{\mathbf{i}}^+ := \min\{t \geq 1 : X_t^\delta = \mathbf{i}\},$$

which we call the first return time to the state  $\mathbf{i}$ .

In order to introduce the second one, it is helpful to visualise the Markov chain  $(X_n^\delta)_{n=1}^\infty$  in the following way. At each transition we append a new bit  $i \in \{1, \dots, N\}$  with probability  $p_i$  on the left of the current state, say  $\mathbf{i} = i_1 \dots i_n$ . There is now a unique way that we can delete the tail of  $i\mathbf{i} = ii_1 \dots i_n$  to yield a new word  $\mathbf{j} = ii_1 \dots i_{n-m} \in P_\delta$ . Then,  $\mathbf{j}$  is the new current state of the chain. In this sense, our Markov chain  $(X_n^\delta)_{n=0}^\infty$  is closely related to the Markov chain which describes observing patterns of heads and tails in coin tossing, and we can exploit this connection to adapt techniques which were used in [9, §11.3.3] to compute waiting times for all patterns of a fixed length when tossing a fair coin.

Let  $w_{\mathbf{i}}$  denote the first time that  $\mathbf{i}$  appears using all new bits, that is, with no overlap with the initial state. This random variable is easier to study than the waiting time  $\tau_{\mathbf{i}}$  since it does not depend on the initial state. There are two trivial observations which will be useful: (i)  $w_{\mathbf{i}} \geq \tau_{\mathbf{i}}$  for all  $\mathbf{i} \in P_\delta$  and (ii) since  $w_{\mathbf{i}}$  does not depend on the initial state,  $\mathbb{E}w_{\mathbf{i}} \geq \mathbb{E}_i \tau_{\mathbf{i}}^+$ , where we have purposefully suppressed the dependence on the initial state in  $\mathbb{E}w_{\mathbf{i}}$ .

Finally, we fix some extra notation which will be used throughout the proofs. We will write  $A \lesssim B$  if  $A \leq cB$  for some constant  $c$  which depends only on the parameters fixed by the hypothesis of the result being proved, which in our case may be the IFS itself and the probability vector  $(p_1, \dots, p_N)$ , but crucially it will never depend on  $\delta$ ,  $v_0$  or  $\mathbf{i}_0$ . Similarly we write  $A \gtrsim B$  if  $B \lesssim A$  and  $A \approx B$  if both  $A \lesssim B$  and  $B \lesssim A$ .

Also recall that in the proof of Proposition 2.3 we introduced the notation

$$L_\delta = \max\{|\mathbf{i}| : \mathbf{i} \in P_\delta\} = \left\lceil \frac{\log \delta}{\log r_{\min}} \right\rceil.$$

Additionally we will denote

$$\ell_\delta = \min\{|\mathbf{i}| : \mathbf{i} \in P_\delta\} = \left\lceil \frac{\log \delta}{\log r_{\max}} \right\rceil.$$

### 3.1. Proofs of upper bounds

First, we obtain the upper bound on  $\mathbb{E}W_{\delta, v_0}$  in the case where (1.1) is not maximised uniquely in  $\{1, \dots, N\}$ , that is, we settle the upper bound in (1.3). For this we will appeal to Matthews’ original upper bound on the expected covering time of an irreducible Markov chain (see [9, Theorem 11.2] based on [11]).

**PROPOSITION 3.1.** *Fix  $0 < \delta < r_{\min}$ . Then*

$$\max_{\mathbf{i} \in P_\delta} \mathbb{E}_i \tau_{\text{cov}} \leq \max_{\mathbf{i}, \mathbf{j} \in P_\delta} \mathbb{E}_i \tau_{\mathbf{j}} \left( 1 + \frac{1}{2} + \dots + \frac{1}{|P_\delta|} \right).$$

We will require the following short lemma which describes which states are most difficult to hit and, given such a state, provides an approximate formula for its stationary probability.

**LEMMA 3.2.** *Let  $i_0 \in \{1, \dots, N\}$  satisfy  $\frac{\log p_{i_0}}{\log r_{i_0}} = \max_i \frac{\log p_i}{\log r_i} = t$ . Given  $0 < \delta < r_{\min}$  let  $\mathbf{i}_0 \in P_\delta$  denote the unique string which is made up only of the digit  $i_0$ . Then*

$$\min_{\mathbf{i} \in P_\delta} p_{\mathbf{i}} \approx p_{\mathbf{i}_0} \approx \delta^t \tag{3.1}$$

where the implied constants are independent of  $\delta$ .



*Proof.* Since  $t = \frac{\log p_{i_0}}{\log r_{i_0}}$ ,

$$p_{i_0} = r_{i_0}^{t|i_0|} = r_{i_0}^t \approx \delta^t, \quad (3.2)$$

where the last approximate equality follows because  $\delta r_{\min} < r_{i_0}^{|i_0|-1} r_{\min} \leq r_{i_0} \leq \delta$  by definition of  $P_\delta$ .

Also, since  $t = \max_i \frac{\log p_i}{\log r_i}$ , for any  $\mathbf{i} = i_1 \dots i_n \in P_\delta$  we have

$$p_{\mathbf{i}} = p_{i_1} \cdots p_{i_n} = r_{i_1}^{\frac{\log p_{i_1}}{\log r_{i_1}}} \cdots r_{i_n}^{\frac{\log p_{i_n}}{\log r_{i_n}}} \geq r_{\mathbf{i}}^t > (\delta r_{\min})^t \approx \delta^t \quad (3.3)$$

where the second inequality follows because  $r_{\mathbf{i}} \geq r_{i_1} \cdots r_{i_{n-1}} r_{\min} > \delta r_{\min}$  by definition of  $P_\delta$ . Combining (3.2) and (3.3) gives (3.1).  $\square$

*Proof of upper bound in (1.3).* Let  $i_0 \in \{1, \dots, N\}$  satisfy  $\frac{\log p_{i_0}}{\log r_{i_0}} = \max_i \frac{\log p_i}{\log r_i} = t$ . Fix arbitrary  $0 < \epsilon < r_{\min}$  throughout the proof, and for each  $k \in \mathbb{N}$  consider  $P_{\epsilon^k}$ . Let  $\mathbf{i} \in P_{\epsilon^k}$ . We can write  $\mathbf{i}$  uniquely as  $\mathbf{i}_k \mathbf{i}_{k-1} \dots \mathbf{i}_1$  where for each  $1 \leq m \leq k$ ,  $\mathbf{i}_m \dots \mathbf{i}_1 \in P_{\epsilon^m}$ . We note that since  $\epsilon < r_{\min}$ ,  $|\mathbf{i}_1| \geq 2$  by definition of  $P_\epsilon$ . Moreover, for each  $1 \leq m \leq k$ ,

$$r_{\mathbf{i}_m \dots \mathbf{i}_1} \leq \epsilon^m < r_{\mathbf{i}_m \dots \mathbf{i}_1 \mathbf{i}'_1}$$

where  $\mathbf{i}'_1$  denotes  $\mathbf{i}_1$  with its last digit removed. Therefore

$$\epsilon^{m+1} < r_{\mathbf{i}_m \dots \mathbf{i}'_1} \epsilon < r_{\mathbf{i}_m \dots \mathbf{i}'_1} r_{\min} \leq r_{\mathbf{i}_m \dots \mathbf{i}_1} \leq \epsilon^m$$

hence  $\mathbf{i}_{m+1}$  must necessarily have non-zero length. Let  $C_1 := \max_{\mathbf{j} \in P_\epsilon} \mathbb{E} w_{\mathbf{j}} < \infty$  and let  $C_2 := \max_{\mathbf{j} \in P_\epsilon} |\mathbf{j}|$ . Let  $C = \max\{C_1, C_2\}$ .

Now, for each  $2 \leq m \leq k$ ,

$$\begin{aligned} \mathbb{E} w_{\mathbf{i}_m \dots \mathbf{i}_1} &\leq (\mathbb{E} w_{\mathbf{i}_{m-1} \dots \mathbf{i}_1} + C) p_{i_m} + (\mathbb{E} w_{\mathbf{i}_{m-1} \dots \mathbf{i}_1} + C + \mathbb{E} w_{\mathbf{i}_m \dots \mathbf{i}_1}) (1 - p_{i_m}) \\ &= \mathbb{E} w_{\mathbf{i}_{m-1} \dots \mathbf{i}_1} + C + (1 - p_{i_m}) \mathbb{E} w_{\mathbf{i}_m \dots \mathbf{i}_1}. \end{aligned} \quad (3.4)$$

Therefore

$$p_{i_m} \mathbb{E} w_{\mathbf{i}_m \dots \mathbf{i}_1} \leq \mathbb{E} w_{\mathbf{i}_{m-1} \dots \mathbf{i}_1} + C. \quad (3.5)$$

In particular, for  $k \geq 2$ ,

$$p_{i_2} \cdots p_{i_k} \mathbb{E} w_{\mathbf{i}_k \dots \mathbf{i}_1} \leq 2C + C p_{i_2} + C p_{i_3 i_2} + \cdots + C p_{i_{k-1} \dots i_2} \leq C + \frac{C}{1-p},$$

where  $p := \max_{i \in \mathcal{I}} p_i$ . Therefore,

$$\mathbb{E} w_{\mathbf{i}} \lesssim p_i^{-1} \quad (3.6)$$

where the implied constant depends on the probability vector  $(p_1, \dots, p_N)$  and on the choice of  $\epsilon$  but does not depend on the length  $k = |\mathbf{i}|$ .

Now, fix  $\delta > 0$ , and let  $k \in \mathbb{N}$  satisfy  $\epsilon^k \leq \delta < \epsilon^{k-1}$ . Let  $\mathbf{j} \in P_\delta$ . Then we can choose some  $\mathbf{i} \in P_{\epsilon^k}$  such that  $\mathbf{i}|_n = \mathbf{j}$  for some  $n \in \mathbb{N}$ . Then  $p_{\mathbf{i}} \geq p_{\mathbf{j}} p_{\min}^{C_2} \approx p_{\mathbf{j}}$  where  $p_{\min} := \min_i p_i$ . In particular

$$\mathbb{E} w_{\mathbf{j}} \leq \mathbb{E} w_{\mathbf{i}} \lesssim p_{\mathbf{i}}^{-1} \lesssim p_{\mathbf{j}}^{-1}. \quad (3.7)$$

Let  $\mathbf{i}_0 \in P_\delta$  be the unique string consisting only of the digit  $i_0$ . By (3.1),

$$\max_{\mathbf{j} \in P_\delta} \mathbb{E} w_{\mathbf{j}} \lesssim p_{\mathbf{j}}^{-1} \lesssim p_{i_0}^{-1} \approx \delta^{-t}. \quad (3.8)$$

By Proposition 3.1,

$$\mathbb{E}_{i_0} \tau_{\text{cov}} \leq \left(1 + \frac{1}{2} + \cdots + \frac{1}{|P_\delta|}\right) \max_{\mathbf{i}, \mathbf{j} \in P_\delta} \mathbb{E}_i \tau_{\mathbf{j}} \leq \left(1 + \frac{1}{2} + \cdots + \frac{1}{N_\delta}\right) \max_{\mathbf{j} \in P_\delta} \mathbb{E} w_{\mathbf{j}} \lesssim \delta^{-t} \log \left(\frac{1}{\delta}\right)$$

where we have used the fact that  $N_\delta := |P_\delta| \leq r_{\min}^{-s} \delta^{-s}$  as remarked at the beginning of the previous section. The result now follows by Proposition 2.1.  $\square$

In the above proof we did not refer to the question of whether or not (1.1) is uniquely maximised only at  $i_0$ , and indeed the upper bound of  $\mathbb{E} W_{\delta, v_0} \lesssim \delta^{-t} \log \left(\frac{1}{\delta}\right)$  for all  $v_0 \in F$  is

valid whether or not this is the case. However, in the case where (1.1) is maximised uniquely this upper bound is not optimal owing to the fact that on a large part of the state space  $P_\delta$ , the expected hitting times are significantly lower than  $\delta^{-t}$ . Indeed for a typical choice of  $\mathbf{i} \in P_\delta$  an expected hitting time  $\mathbb{E}_i \tau_j$  will be of the order  $\delta^{-t}$  only if the word  $\mathbf{j}$  contains many instances of the digit  $i_0$ . Thus in the case where (1.1) is maximised at a unique index it is useful to separate the less accessible part of the state space from the more accessible part of the state space and estimate the expected covering times of each of these parts separately.

To be precise, for any  $B \subset P_\delta$  we define

$$\tau_{\text{cov}}^B = \min\{t \geq 0 : \forall \mathbf{i} \in B, \exists s \leq t \text{ s.t. } X_s = \mathbf{i}\},$$

that is, the first time that all of the states in  $B$  have been visited by the chain. The expected covering time of  $B$  satisfies an analogous upper bound to the expected covering time (see [9, (11.16)] which can easily be proven by adapting the proof of [9, Theorem 11.2]):

PROPOSITION 3.3. *Fix  $0 < \delta < r_{\min}$  and  $B \subset P_\delta$ . Then*

$$\max_{\mathbf{i} \in B} \mathbb{E}_i \tau_{\text{cov}}^B \leq \max_{\mathbf{i}, \mathbf{j} \in B} \mathbb{E}_i \tau_j \left(1 + \frac{1}{2} + \cdots + \frac{1}{|B|}\right).$$

In particular, in the case where (1.1) is uniquely maximised one can improve on the upper bound of  $\mathbb{E}W_{\delta, v_0} \lesssim \delta^{-t} \log\left(\frac{1}{\delta}\right)$  by instead estimating how long it would take for the Markov chain to first visit all states in a subset  $B$  consisting of states  $\mathbf{i} \in P_\delta$  which contain a restricted number of the digit  $i_0$ , followed by visiting all states in  $B' = P_\delta \setminus B$ . The subset  $B$  would constitute most of the state space  $P_\delta$ , but would benefit from having a reduced upper bound on  $\max_{\mathbf{i}, \mathbf{j} \in B} \mathbb{E}_i \tau_j$ . On the other hand, the upper bound on  $\max_{\mathbf{i}, \mathbf{j} \in B'} \mathbb{E}_i \tau_j$  would be of the order  $\delta^{-t}$ , but  $B'$  would only comprise a small proportion of the state space. Of course, there is a lot of flexibility in how the subset  $B$  could be defined, so by carefully considering the contribution of each covering time  $\mathbb{E}_i \tau_{\text{cov}}^B$  and  $\mathbb{E}_i \tau_{\text{cov}}^{B'}$  one can choose  $B$  in such a way that the upper bound on  $\mathbb{E}W_{\delta, v_0}$  is improved to the degree noted in Theorem 1.1:

*Proof of upper bound in (1.2).* Suppose that  $i_0 \in \{1, \dots, N\}$  is the unique digit that satisfies  $\frac{\log p_{i_0}}{\log r_{i_0}} = \max_i \frac{\log p_i}{\log r_i} = t$ . Fix  $\delta > 0$  and consider  $k_\delta$  satisfying  $1 \leq k_\delta \leq \ell_\delta$  to be chosen later. Define  $B \subset P_\delta$  to be all strings which contain at most  $k_\delta - 1$  digits from the set  $\{1, \dots, N\} \setminus \{i_0\}$ . Also denote  $B' = P_\delta \setminus B$ , so that  $B'$  are all strings in  $P_\delta$  which contain at least  $k_\delta$  digits from the set  $\{1, \dots, N\} \setminus \{i_0\}$ .

We can bound the covering time above by the time it would take to first cover  $B$  and then cover  $B'$  (plus the intermediate travel time) yielding

$$\max_{\mathbf{i} \in P_\delta} \mathbb{E}_i \tau_{\text{cov}} \leq \max_{\mathbf{i} \in P_\delta} \min_{\mathbf{j} \in B} \mathbb{E}_i \tau_j + \max_{\mathbf{i} \in B} \mathbb{E}_i \tau_{\text{cov}}^B + \max_{\mathbf{i} \in P_\delta} \min_{\mathbf{j} \in B'} \mathbb{E}_i \tau_j + \max_{\mathbf{i} \in B'} \mathbb{E}_i \tau_{\text{cov}}^{B'}. \quad (3.9)$$

We will use Proposition 3.3 to bound  $\max_{\mathbf{i} \in B} \mathbb{E}_i \tau_{\text{cov}}^B$  and  $\max_{\mathbf{i} \in B'} \mathbb{E}_i \tau_{\text{cov}}^{B'}$  in terms of  $k_\delta$ , before choosing  $k_\delta$  in a way that optimises this bound.

To calculate  $|B|$ , notice that  $B = \bigcup_{m=0}^{k_\delta-1} B_m$  where  $B_m$  is the set of strings in  $P_\delta$  which contain exactly  $m$  occurrences of digits from the set  $\{1, \dots, N\} \setminus \{i_0\}$ . Since any  $\mathbf{i} \in B$  has length  $|\mathbf{i}| \leq L_\delta$ , there are no more than  $L_\delta$  positions where the first digit from the set  $\{1, \dots, N\} \setminus \{i_0\}$  can appear, followed by no more than  $L_\delta - 1$  positions where the second digit from the set  $\{1, \dots, N\} \setminus \{i_0\}$  can appear and so on. Since we have  $N - 1$  possible choices of digits for each of these,

$$|B_m| \leq (N - 1)^m L_\delta (L_\delta - 1) \cdots (L_\delta - m + 1) \leq (N - 1)^m L_\delta^m.$$

Therefore,

$$|B| \leq \sum_{m=0}^{k_\delta-1} (N - 1)^m L_\delta^m \lesssim ((N - 1)L_\delta)^{k_\delta}.$$

Since  $L_\delta \lesssim \log\left(\frac{1}{\delta}\right)$  we have  $\log|B| \lesssim k_\delta \log \log\left(\frac{1}{\delta}\right)$ . Notice that  $B$  contains a string  $\mathbf{i}_0$  which contains only the digit  $i_0$ , therefore  $\min_{\mathbf{i} \in B} p_i \approx p_{\mathbf{i}_0} \approx \delta^t$  by (3.1).

Next we consider  $B'$ . Since  $B' = P_\delta \setminus B$  we have  $\log|B'| \approx \log\left(\frac{1}{\delta}\right)$ . Next we calculate  $\min_{\mathbf{i} \in B'} p_i$ . Let  $\mathbf{i} \in B'$ . Then for some integer  $m$  satisfying  $k_\delta \leq m \leq L_\delta$ ,  $\mathbf{i}$  contains  $m$  digits

from the set  $\{1, \dots, N\} \setminus \{i_0\}$ . In particular there exist  $j_1, \dots, j_m \in \mathcal{I} \setminus \{i_0\}$  and some integer  $n$  such that

$$r_i = r_{j_1} \cdots r_{j_m} r_{i_0}^n \approx \delta.$$

We have

$$p_{i_0}^n = r_{i_0}^{\frac{n \log p_{i_0}}{\log r_{i_0}}} \approx \left( \frac{\delta}{r_{j_1} \cdots r_{j_m}} \right)^t.$$

Therefore

$$p_i \approx \delta^t \frac{p_{j_1} \cdots p_{j_m}}{r_{j_1}^t \cdots r_{j_m}^t}.$$

Fix

$$c = t - \max_{i \neq i_0} \frac{\log p_i}{\log r_i} > 0,$$

where positivity follows from the fact that  $i_0$  uniquely achieves the maximum in (1.1). It follows that for each  $i \in \{1, \dots, N\} \setminus \{i_0\}$ ,

$$\frac{p_i}{r_i^t} = \frac{r_i^{\frac{\log p_i}{\log r_i}}}{r_i^t} = r_i^{\frac{\log p_i}{\log r_i} - t} \geq r_i^{-c}.$$

Therefore

$$p_i \gtrsim \delta^t r_i^{-cm} \geq \delta^t r_{\max}^{-ck_\delta}.$$

Note that by (3.7),

$$\max_{i \in P_\delta} \min_{j \in B} \mathbb{E}_i \tau_j \leq \min_{j \in B} \mathbb{E} w_j \lesssim \min_{j \in B} p_j^{-1} \lesssim \delta^{-t}$$

and

$$\max_{i \in P_\delta} \min_{j \in B'} \mathbb{E}_i \tau_j \leq \min_{j \in B'} \mathbb{E} w_j \lesssim \min_{j \in B'} p_j^{-1} \lesssim \delta^{-t} r_{\max}^{ck_\delta}.$$

Therefore, by Proposition 3.3 and (3.9),

$$\max_{i \in P_\delta} \mathbb{E}_i \tau_{\text{cov}} \lesssim \delta^{-t} k_\delta \log \log \left( \frac{1}{\delta} \right) + \delta^{-t} r_{\max}^{ck_\delta} \log \left( \frac{1}{\delta} \right). \quad (3.10)$$

Consider the function  $f_\delta(x) = x \log \log \left( \frac{1}{\delta} \right) - r_{\max}^{ck_\delta} \log \left( \frac{1}{\delta} \right)$ . Observe that  $f_\delta(2) < 0$  and  $f_\delta(\ell_\delta) > 0$ , provided  $\delta$  is sufficiently small. Therefore there exists  $2 < x_\delta < \ell_\delta$  such that  $f_\delta(x_\delta) = 0$ . Notice that  $f_\delta(\log \log \left( \frac{1}{\delta} \right)) < 0$  provided  $\delta$  is sufficiently small. Since  $f_\delta(x)$  is increasing with  $x$ ,  $x_\delta > \log \log \left( \frac{1}{\delta} \right)$ . Therefore, if we fix  $k_\delta = \lfloor x_\delta \rfloor$ , we have

$$k_\delta \log \log \frac{1}{\delta} \approx r_{\max}^{ck_\delta} \log \left( \frac{1}{\delta} \right)$$

and  $k_\delta \lesssim \log \log \left( \frac{1}{\delta} \right)$ . Hence by (3.10),

$$\max_{i_0 \in P_\delta} \mathbb{E}_{i_0} \tau_{\text{cov}} \lesssim \delta^{-t} k_\delta \log \log \left( \frac{1}{\delta} \right) \leq \delta^{-t} \left( \log \log \left( \frac{1}{\delta} \right) \right)^2.$$

The result now follows from Proposition 2.1.  $\square$

### 3.2. Proofs of lower bounds

To estimate the lower bounds on the expected covering time, we begin by obtaining a lower estimate for the expected hitting time  $\mathbb{E}_i \tau_j$  in terms of  $\mathbb{E} w_j$ .

LEMMA 3.4. *Given arbitrary  $i, j \in P_\delta$  where  $i \neq j$  denote*

$$\theta_{i,j} = \mathbb{P}_i(\tau_j < L_\delta).$$

Then

$$\mathbb{E}_i \tau_j \geq \mathbb{E} w_j - \theta_{i,j}(L_\delta + \mathbb{E} w_j). \quad (3.11)$$

*Proof.* Observe that

$$w_j \leq \tau_j + \mathbb{1}_{\{\tau_j < L_\delta\}}(L_\delta + w_j^*) \quad (3.12)$$

where  $w_j^*$  is the amount of time required to build  $j$  from new bits after the  $L_\delta$ th bit has been added. Indeed (3.12) holds since if  $(X_t)$  is such that  $j$  appears for the first time after  $L_\delta$  time then  $w_j((X_t)) = \tau_j((X_t))$  whereas if  $(X_t)$  is such that  $j$  appears for the first time before  $L_\delta$  time then  $w_j((X_t)) \leq L_\delta + w_j^*((X_t))$ .

Now, since  $w_j^*$  is independent of the event  $\{\tau_j < L_\delta\}$  and  $w_j$  has the same distribution as  $w_j^*$ , we can take expectations in (3.12) to obtain

$$\mathbb{E}w_j \leq \mathbb{E}_i \tau_j + \theta_{i,j}(L_\delta + \mathbb{E}w_j)$$

which completes the proof of (3.11).  $\square$

The usefulness of (3.11) is that  $\mathbb{E}w_j \geq \mathbb{E}_j \tau_j^+$ , and the expected return time  $\mathbb{E}_j \tau_j^+$  satisfies the following formula (see [9, Proposition 1.14]).

PROPOSITION 3.5. Fix  $0 < \delta < r_{\min}$ . For all  $i \in P_\delta$ ,

$$\mathbb{E}_i(\tau_i^+) = \frac{1}{\pi_i} = \frac{1}{p_i}.$$

So, in order to apply (3.11) we require an estimate on the probability  $\theta_{i,j}$  of a fast hitting time of state  $j$  from state  $i$ , which is provided by the following lemma.

LEMMA 3.6. Fix  $i, j \in P_\delta$ , where  $i \neq j$ . Suppose that  $\tau_j((X_t)) \geq j$  whenever  $X_0 = i$ . Then

$$\theta_{i,j} = \mathbb{P}_i(\tau_j < L_\delta) \leq \frac{p^j}{1-p} + p^{\ell_\delta}(L_\delta - \ell_\delta).$$

*Proof.* Fix such  $i$  and  $j$ . We have

$$\mathbb{P}_i(\tau_j < L_\delta) \leq \mathbb{P}_i(\tau_j = j) + \mathbb{P}_i(\tau_j = j+1) + \cdots + \mathbb{P}_i(\tau_j = L_\delta - 1).$$

For each  $j \leq k \leq |j| - 1$ ,  $\mathbb{P}_i(\tau_j = k) \leq p^k$ , since  $k$  correct transitions are required (which correspond to the  $k$  correct digits that need to be appended to the left of the word  $i$ ). If  $|j| < L_\delta$ , for each  $|j| \leq k \leq L_\delta - 1$ ,  $\mathbb{P}_i(\tau_j = k) \leq p^{|j|}$ , since  $|j|$  correct transitions are required. Therefore

$$\begin{aligned} \mathbb{P}_i(\tau_j < L_\delta) &\leq \mathbb{P}_i(\tau_j = j) + \mathbb{P}_i(\tau_j = j+1) + \cdots + \mathbb{P}_i(\tau_j = L_\delta - 1) \\ &\leq p^j + p^{j+1} + \cdots + p^{|j|-1} + (L_\delta - |j|)p^{|j|} \\ &\leq p^j + p^{j+1} + \cdots + p^{\ell_\delta-1} + (L_\delta - \ell_\delta)p^{\ell_\delta} \\ &= p^j \frac{1 - p^{\ell_\delta-j-2}}{1-p} + (L_\delta - \ell_\delta)p^{\ell_\delta} \\ &\leq \frac{p^j}{1-p} + p^{\ell_\delta}(L_\delta - \ell_\delta). \end{aligned}$$

$\square$

We are almost ready to prove the lower bounds on  $\mathbb{E}W_{\delta, v_0}$  from (1.2) and (1.3) via appropriate lower bounds on the expected covering time of  $(X_n^{\delta})_{n=0}^{\infty}$ . Analogously to Proposition 3.1, the original lower bound of Matthews [11] bounds the minimum expected covering time from below by the minimum expected hitting times between distinct states multiplied by the logarithm of the cardinality of the state space. Clearly, this bound is insufficient for our purposes, since it can yield a lower bound merely of the order  $\log(\frac{1}{\delta})$  owing to the fact in general some states in  $P_\delta$  will be extremely close to one other. Instead, we can again improve on this bound by considering the expected covering time of a subset of the state space, where this time the elements in the subset are chosen in such a way that they are all “uniformly far” from each other in the sense

that the expected hitting times of possible pairs of states are uniformly bounded below. For this we will require the following analogue of Proposition 3.3 (see [9, Proposition 11.4]<sup>†</sup>).

PROPOSITION 3.7. *Let  $0 < \delta < r_{\min}$  and  $B \subset P_\delta$ . Then for all  $i_0 \in B$ ,*

$$\mathbb{E}_{i_0} \tau_{\text{cov}} \geq \mathbb{E}_{i_0} \tau_{\text{cov}}^B \geq \min_{i,j \in B, i \neq j} \mathbb{E}_i(\tau_j) \left( 1 + \frac{1}{2} + \cdots + \frac{1}{|B|-1} \right).$$

We are now ready to obtain a lower bound on  $\mathbb{E}W_{\delta, v_0}$  in the case that (1.1) is maximised uniquely at some  $i_0 \in \{1, \dots, N\}$ . Since the least accessible part of the state space  $P_\delta$  comprises states  $i \in P_\delta$  which consist mostly of the digit  $i_0$ , the most effective choice of  $B$  in Proposition 3.7 is a subset of this type. Then applying the lower estimates from Lemma 3.4 on the expected hitting time will yield the desired result.

*Proof of lower bound in (1.2).* Fix  $\delta > 0$  sufficiently small such that  $\frac{p^{\ell_\delta - 1}}{1-p} + p^{\ell_\delta}(L_\delta - \ell_\delta) \leq \frac{1}{2}$  and denote  $i_0 \in \{1, \dots, N\}$  to be the digit that satisfies  $\frac{\log p_{i_0}}{\log r_{i_0}} = \max_i \frac{\log p_i}{\log r_i} = t$ . Let  $j \in \{1, \dots, N\} \setminus \{i_0\}$ . Define

$$A = \{i \in P_{\delta/r_j^2} : i \text{ contains one instance of } j \text{ and } |i| - 1 \text{ instances of } i_0\}$$

and  $B = \{jji : i \in A\} \subset P_\delta$ . For any  $j \in B$ ,  $\pi_j = p_j \approx \delta^t$  and therefore for any  $i, j \in B$ , where  $i \neq j$  we have

$$\mathbb{E}_i \tau_j \geq (1 - \theta_{i,j}) \mathbb{E}w_j - L_\delta \theta_{i,j} \geq (1 - \theta_{i,j}) \mathbb{E}_j \tau_j^+ - L_\delta \theta_{i,j} \gtrsim (1 - \theta_{i,j}) \delta^{-t} - L_\delta \theta_{i,j}$$

by (3.11) and Proposition 3.5. In order to bound  $\theta_{i,j}$ , notice that by definition of  $B$ , at least  $|j| - 1$  transitions are required to hit the state  $j$  from  $i$ . Thus by Lemma 3.6 and our assumption on  $\delta$ ,

$$\theta_{i,j} \leq \frac{p^{|j|-1}}{1-p} + p^{\ell_\delta}(L_\delta - \ell_\delta) \leq \frac{p^{\ell_\delta - 1}}{1-p} + p^{\ell_\delta}(L_\delta - \ell_\delta) \leq \frac{1}{2}.$$

Finally, to calculate  $|B|$ , observe that there are at least  $l_{\delta/r_j^2} \approx \log \frac{1}{\delta}$  distinct positions at which the digit  $j$  can be placed within a string  $i \in A$  and therefore  $\log |B| = \log |A| \approx \log \log \left( \frac{1}{\delta} \right)$ .

By Proposition 3.7

$$\begin{aligned} \min_{i_0 \in P_\delta} \mathbb{E}_{i_0} \tau_{\text{cov}} &\geq \min_{i \in P_\delta} \min_{j \in B} \mathbb{E}_i \tau_j + \min_{i \in B} \mathbb{E}_i \tau_{\text{cov}}^B \\ &\geq \left( 1 + \frac{1}{2} + \cdots + \frac{1}{|B|-1} \right) \min_{\substack{i,j \in B \\ i \neq j}} \mathbb{E}_i \tau_j \\ &\gtrsim \delta^{-t} \log \log \left( \frac{1}{\delta} \right). \end{aligned}$$

The result follows by Proposition 2.1. □

All that remains is for us to obtain the lower bound in Theorem 1.1 in the case that (1.1) is not uniquely maximised in  $\{1, \dots, N\}$ . Since there must be at least two digits which attain this maximum, Proposition 3.7 can be applied for a choice of  $B \subset P_\delta$  where the cardinality of  $B$  is exponential in  $\delta^{-1}$ . This allows us to recover a sharp lower bound for  $\mathbb{E}W_{\delta, v_0}$  in this case.

*Proof of lower bound in (1.3).* Let  $\mathcal{J} \subset \{1, \dots, N\}$  be the set for which the maximum in (1.1) is attained, where  $|\mathcal{J}| \geq 2$  by assumption. Define

$$P'_\delta = P_\delta \cap \mathcal{J}^*$$

where  $\mathcal{J}^*$  denotes the set of finite words with digits in  $\mathcal{J}$ . We begin by showing that there exists  $a > 0$  such that  $|P'_\delta| \approx \delta^{-a}$ . Let  $a > 0$  satisfy  $\sum_{i \in \mathcal{J}} r_i^a = 1$  and let  $\mathbb{P}$  be the Bernoulli measure on

---

<sup>†</sup> Although Proposition 11.4 in [9] is not stated exactly as it is here, Proposition 3.7 can easily be gleaned from the proof of [9, Proposition 11.4].

$\Sigma$  where  $\mathbb{P}([i]) = r_i^a$  if  $i \in \mathcal{J}$  and  $\mathbb{P}[i] = 0$  otherwise. Then  $1 = \sum_{i \in P_\delta} p_i = \sum_{i \in P'_\delta} r_i^a$ , therefore  $r_{\min}^a \delta^a |P'_\delta| < 1 \leq \delta^a |P'_\delta|$ , implying that  $|P'_\delta| \approx \delta^{-a}$ .

Fix  $j \in \mathbb{N}$  sufficiently large such that  $\frac{p^j}{1-p} \leq \frac{1}{4}$ . Fix  $\delta > 0$  sufficiently small such that  $\ell_\delta > j$  and  $p^{\ell_\delta} (L_\delta - \ell_\delta) \leq \frac{1}{4}$ . Fix arbitrary  $\mathbf{i}_0 \in P_\delta$  and denote the first  $j$  digits of  $\mathbf{i}_0$  by  $i_1 \dots i_j$ . Define a new word  $\mathbf{k} = k_1 \dots k_j$  by setting  $k_1 = i_1$  and fixing  $k_2 = \dots = k_j = w_1 \in \mathcal{J} \setminus \{i_1\}$ . Note that it is not necessarily true that  $\{i_1\} \subset \mathcal{J}$ , but if this is the case then  $w_1 \in \mathcal{J} \setminus \{i_1\}$  can always be chosen since  $|\mathcal{J}| \geq 2$ . Define  $B_{\mathbf{i}_0} \subset P_\delta$  as the set

$$B_{\mathbf{i}_0} = \{\mathbf{i} \in P_\delta : \mathbf{i} = \mathbf{k}\mathbf{j} \text{ for some } \mathbf{j} \in \mathcal{J}^*\}.$$

We begin by claiming that for all  $\delta > 0$  sufficiently small and all  $\mathbf{i} \in B_{\mathbf{i}_0}$ ,  $\pi_{\mathbf{i}} \lesssim \delta^t$ . Writing  $\mathbf{i} = \mathbf{k}\mathbf{j}$  as in the definition of  $B_{\mathbf{i}_0}$ , we have

$$\pi_{\mathbf{i}} = p_{\mathbf{i}} = p_{\mathbf{k}} p_{\mathbf{j}} \leq \frac{\delta^t}{r_{\min}^{jt}} p^j \lesssim \delta^t.$$

Next we estimate  $\theta_{\mathbf{i},\mathbf{j}} = \mathbb{P}_{\mathbf{i}}(\tau_{\mathbf{j}}^+ < L_\delta)$  for  $\mathbf{i}, \mathbf{j} \in B_{\mathbf{i}_0}$ ,  $\mathbf{i} \neq \mathbf{j}$ . Fix such  $\mathbf{i}$  and  $\mathbf{j}$ . Observe that both  $\mathbf{i}|_j = \mathbf{k}$  and  $\mathbf{j}|_j = \mathbf{k}$ . Since  $k_1$  does not agree with  $k_2, \dots, k_j$ , at least  $j$  transitions are required before the chain can hit the state  $\mathbf{j}$ , when starting from state  $\mathbf{i}$ . Therefore by Lemma 3.6 and our assumptions on  $j$  and  $\delta$ ,

$$\theta_{\mathbf{i},\mathbf{j}} \leq \frac{p^j}{1-p} + (L_\delta - \ell_\delta) p^{\ell_\delta} \leq \frac{1}{2}.$$

Next we bound  $|B_{\mathbf{i}_0}|$ . Observe that

$$B_{\mathbf{i}_0} = \{\mathbf{k}\mathbf{j} : \mathbf{j} \in P'_{\frac{\delta}{r_{\mathbf{k}}}}\},$$

therefore  $|B_{\mathbf{i}_0}| = |P'_{\frac{\delta}{r_{\mathbf{k}}}}| \approx \delta^{-a}$ , where  $a$  satisfies  $\sum_{i \in \mathcal{J}} r_i^a = 1$ , as before. Therefore by Lemma 3.7, for any  $\mathbf{j} \in B_{\mathbf{i}_0}$ ,

$$\mathbb{E}_{\mathbf{j}}(\tau_{\text{cov}}^{B_{\mathbf{i}_0}}) \gtrsim \delta^{-t} \log\left(\frac{1}{\delta}\right).$$

By Proposition 3.7,

$$\min_{\mathbf{i}_0 \in P_\delta} \mathbb{E}_{\mathbf{i}_0} \tau_{\text{cov}} \geq \min_{\mathbf{i}_0 \in P_\delta} \min_{\mathbf{j} \in B_{\mathbf{i}_0}} \mathbb{E}_{\mathbf{i}_0}(\tau_{\mathbf{j}}) + \min_{\mathbf{i}_0 \in P_\delta} \min_{\mathbf{j} \in B_{\mathbf{i}_0}} \mathbb{E}_{\mathbf{j}}(\tau_{\text{cov}}^{B_{\mathbf{i}_0}}) \gtrsim \delta^{-t} \log\left(\frac{1}{\delta}\right).$$

The result follows by Proposition 2.1.  $\square$

#### 4. Directions for future research

Besides the problem of obtaining a sharp estimate for the asymptotic behaviour of the expected  $\delta$ -waiting time in the case where the maximum in (1.1) is attained uniquely, several further directions of research suggest themselves. On the one hand, while this work helps to shed light on how the sequence  $(x_n)_{n=0}^\infty$  approaches the attractor set, we have not investigated the related question of how quickly the measures  $\frac{1}{n} \sum_{k=0}^{n-1} \delta_{x_k}$  approach the self-similar limit measure  $m = \sum_{i=1}^N r_i(S_i) * m$  (with respect to, for example, the Wasserstein distance) and this question may be of interest in future research. It is also interesting to ask how far these results may be extended to the context of iterated function systems defined by maps which are not similarities (such as affine or conformal differentiable transformations) and to cases where the open set condition is not satisfied. Finally, we note that there are analogous questions which make sense for deterministic chaotic dynamical systems. For example, if  $T: \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$  is the doubling map  $T(x) := 2x \pmod{1}$ , then for Lebesgue almost every  $x \in \mathbb{R}/\mathbb{Z}$  the sequence  $\{x, Tx, T^2x, \dots\}$  is dense in  $\mathbb{R}/\mathbb{Z}$ . One could just as easily ask how the expectation with respect to  $x$  of the first integer  $n$  such that the sequence  $\{x, Tx, \dots, T^{n-1}x\}$  is  $\delta$ -dense in  $\mathbb{R}/\mathbb{Z}$  behaves as a function of  $\delta$  in the limit  $\delta \rightarrow 0$ . For the doubling map  $T(x) := 2x \pmod{1}$  this question can be reduced via Markov partitions to the coupon-collector's problem, but for smooth expanding maps or even Anosov diffeomorphisms the details of such an argument are less clear.

## 5. Acknowledgements

Both authors were financially supported by the *Leverhulme Trust* (Research Project Grant number RPG-2016-194). NJ was also financially supported by the *EPSRC* (Standard Grant EP/R015104/1). The authors thank Kenneth Falconer for suggesting useful references. NJ thanks John Sylvester for enlightening conversations about covering problems for Markov chains.

## References

1. BÁRÁNY, B., HOCHMAN, M., AND RAPAPORT, A. Hausdorff dimension of planar self-affine sets and measures. *Invent Math.* 216, 3 (2019), 601–659.
2. BARNESLEY, M. *Fractals everywhere*. Academic Press, Inc., Boston, MA, 1988.
3. BARNESLEY, M. F., AND DEMKO, S. Iterated function systems and the global construction of fractals. *Proc. Roy. Soc. London Ser. A* 399, 1817 (1985), 243–275.
4. DAS, T., AND SIMMONS, D. The Hausdorff and dynamical dimensions of self-affine sponges: a dimension gap result. *Invent. Math.* 210, 1 (2017), 85–134.
5. ELTON, J. H. An ergodic theorem for iterated maps. *Ergodic Theory Dynam. Systems* 7, 4 (1987), 481–488.
6. GUTIÉRREZ, J. M., IGLESIAS, A., AND RODRÍGUEZ, M. A. A multifractal analysis of IFSP invariant measures with application to fractal image generation. *Fractals* 4, 1 (1996), 17–27.
7. HOCHMAN, M. On self-similar sets with overlaps and inverse theorems for entropy. *Ann. of Math. (2)* 180, 2 (2014), 773–822.
8. HUTCHINSON, J. E. Fractals and self-similarity. *Indiana Univ. Math. J.* 30, 5 (1981), 713–747.
9. LEVIN, D. A., AND PERES, Y. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2017.
10. LINDENSTRAUSS, E., AND VARJÚ, P. P. Random walks in the group of Euclidean isometries and self-similar measures. *Duke Math. J.* 165, 6 (2016), 1061–1127.
11. MATTHEWS, P. Covering problems for Markov chains. *Ann. Probab.* 16, 3 (1988), 1215–1228.
12. SCHIEF, A. Separation properties for self-similar sets. *Proc. Amer. Math. Soc.* 122, 1 (1994), 111–115.
13. SHMERKIN, P. On Furstenberg’s intersection conjecture, self-similar measures, and the  $L^q$  norms of convolutions. *Ann. of Math. (2)* 189, 2 (2019), 319–391.
14. SIMMONS, D., AND WEISS, B. Random walks on homogeneous spaces and Diophantine approximation on fractals. *Invent. Math.* 216, 2 (2019), 337–394.
15. WU, M. A proof of Furstenberg’s conjecture on the intersections of  $\times p$ - and  $\times q$ -invariant sets. *Ann. of Math. (2)* 189, 3 (2019), 707–751.

Ian D. Morris  
 School of Mathematical Sciences, Queen  
 Mary, University of London, Mile End  
 Road, London E1 4NS  
 United Kingdom  
 i.morris@qmul.ac.uk

Natalia Jurga  
 Mathematical Institute, University of St  
 Andrews, KY16 9SS  
 Scotland  
 naj1@st-andrews.ac.uk