

## Dyadic Diophantine approximation and Katok's horseshoe approximation

by

TOMAS PERSSON AND JÖRG SCHMELING (Lund)

**1. Introduction.** In this paper we deal with two questions which are connected to a general principle in dynamical systems theory. This principle is the asymptotics of hitting times. By this we mean the following. Given a dynamical system  $f: M \rightarrow M$  on some compact metric space  $M$  we specify a point  $y \in M$ . We are interested in the set of points which meet a shrinking base of neighbourhoods  $C_n(y)$  of this point with a given speed. More precisely, let  $\xi$  be a finite sufficiently regular partition of  $M$ . We consider cylinder sets  $C_n(y) \in \bigvee_{k=0}^{n-1} f^{-k}\xi$  containing  $y$ . The first hitting time of  $x \in M$  is defined as  $\tau_n(x, y) := \inf\{k \in \mathbb{N} : f^k(x) \in C_n(y)\}$ . The quantity we are interested in is the lower asymptotics of the first hitting time when  $n$  tends to infinity. Using results obtained earlier we are going to apply this setup to two different questions. The first one is purely number-theoretical. The approximation of a real number by rational numbers ordered with respect to their denominator is a well-studied and important problem in number theory. We are going to ask what happens if one does not allow all rationals to approximate a real number but rather rationals with denominators being powers of a given number. In this paper we choose this number to be 2. We remark that there is nothing special in this case. The proofs and statements will be the same for any other natural number. There is a major difference in comparison to the standard Diophantine approximation. While the underlying dynamical system (for a fixed irrational number) in the standard case is an irrational rotation, the underlying case in the dyadic approximation is multiplication by 2. These dynamical systems exhibit a completely different behaviour. Irrational rotations are uniquely ergodic and of zero entropy while multiplication by 2 has a very rich space of invariant measures. Moreover, it has positive topological entropy. Therefore one can expect a

---

2000 *Mathematics Subject Classification*: Primary 11J83; Secondary 37A45, 37C45.

*Key words and phrases*: Diophantine approximation, non-uniformly hyperbolic systems, horseshoes,  $\beta$ -shifts.

more complex and even irregular behaviour of the dyadic approximation than of the standard approximation. These differences are discussed in the corresponding sections.

The second question is more of a dynamical origin although it has a clear interpretation in the expansions of a real number with respect to a non-integer basis ( $\beta$ -expansions). To a real number  $\beta > 1$  one can associate the dynamical system  $x \mapsto \beta x \pmod{1}$ . If one considers the partition of  $[0, 1]$  corresponding to the different branches of this piecewise linear map one obtains a coding space, the  $\beta$ -shift  $S_\beta$ . For some  $\beta$ 's these shifts are especially simple, namely subshifts of finite type. For most numbers these shifts are more complicated. In any case Katok's horseshoe theorem [8] states that these shifts can be approximated from inside by subshifts of finite type. Here we want to make some remarks on this. First, the theory of subshifts of finite type is well understood. In particular, one has control over the statistics of most orbits. This control depends on the size of the transition matrix defining this subshift. So one can expect to have a good control over the statistics of general subshifts if one has control over the speed of approximations by finite type subshifts. This is the question we address in the second part of the paper.

From the methods used in this paper, we arrive at a general dimension transference principle, that dimensions of subsets of arbitrary  $\beta$ -shifts can be calculated as limits of subsets of  $\beta$ -shifts of finite type. This is a useful result, since dimensions are much easier to calculate in subshifts of finite type than in arbitrary subshifts.

**2. Preliminaries.** For simpler notations and formulae we will only use logarithms to base 2 throughout this paper.

**2.1. Dimension.** Let  $Y$  be a subset of  $\mathbb{R}^n$ . Let  $N(\varepsilon)$  denote the minimal number of  $\varepsilon$  balls needed to cover  $Y$ .

DEFINITION 1. For a subset  $Y$  of  $\mathbb{R}^n$ , the *lower box dimension* of  $Y$ , denoted by  $\dim_{\text{LB}} Y$ , is given by

$$\liminf_{\varepsilon \rightarrow 0} \frac{\log N(\varepsilon)}{\log 1/\varepsilon}.$$

The *upper box dimension*  $\dim_{\text{UB}} Y$  is defined similarly, with the  $\liminf$  replaced by  $\limsup$ . If  $\dim_{\text{UB}} Y$  and  $\dim_{\text{LB}} Y$  both exist and are equal, we define the *box dimension* of  $Y$  to be this value, and write  $\dim_{\text{B}} Y = \dim_{\text{UB}} Y = \dim_{\text{LB}} Y$ .

For a subset  $U$  of  $\mathbb{R}^n$ , we let  $\text{diam}(U)$  denote the diameter of  $U$ .

DEFINITION 2. Let  $s \in [0, \infty]$ . The *s-dimensional Hausdorff measure*  $\mathcal{H}^s(Y)$  of a subset  $Y$  of  $\mathbb{R}^n$  is defined by the following limit of covering

sums:

$$\mathcal{H}^s(Y) = \lim_{\varepsilon \rightarrow 0} \left( \inf \left\{ \sum_{i=1}^{\infty} (\text{diam } U_i)^s : Y \subset \bigcup_{i=1}^{\infty} U_i, \sup_i \text{diam } U_i \leq \varepsilon \right\} \right).$$

It is easy to see that there exists a unique  $s_0 = s_0(Y)$  such that

$$(1) \quad \mathcal{H}^s(Y) = \begin{cases} \infty & \text{for } s < s_0, \\ 0 & \text{for } s > s_0. \end{cases}$$

DEFINITION 3. The unique number  $s_0$  given by (1) is defined to be the Hausdorff dimension of  $Y$  and is denoted by  $\dim_{\text{H}} Y$ .

Standard arguments show that for a subset  $Y$  of  $\mathbb{R}^n$ ,

$$\dim_{\text{H}} Y \leq \dim_{\text{LB}} Y \leq \dim_{\text{UB}} Y.$$

There are examples which show that these inequalities may be strict.

The box dimension can also be defined in terms of covering sums, the only change being that the covering sets all have equal diameter. We note that in order to estimate the box dimension, it suffices that the diameters of the covering sets tend to 0 along a geometric sequence.

Lastly, we define the Hausdorff dimension of a measure:

DEFINITION 4. Let  $\mu$  be a Borel probability measure on  $X$ . Then the Hausdorff dimension of the measure  $\mu$  is defined by

$$\dim_{\text{H}} \mu = \inf_Y \{ \dim_{\text{H}} Y : \mu(Y) = 1 \}.$$

We remark that the dimension of a measure is clearly always less than or equal to the dimension of its (Borel) support. There is a well known method of computing lower bounds on the dimension of a measure or a set. Let  $\mathcal{U}(x, r)$  denote the ball of radius  $r$  centred at  $x$ .

LEMMA 1. *If for some finite measure  $\mu$ ,*

$$\liminf_{r \rightarrow 0} \frac{\log \mu(\mathcal{U}(x, r))}{\log r} \geq s \quad \text{on a set of } \mu\text{-positive measure,}$$

*then the dimension of the measure is at least  $s$ .*

A survey of the methods and results in dimension theory can be found in [4, 11].

**2.2. Entropy.** The notion of topological entropy for non-compact sets was introduced by Bowen in [2]. Later it was considered by Pesin and Pitskel' in [12].

The main idea is to replace the diameter as a geometric measure by a measurement based on the complexity. Roughly speaking, we are going to measure our system with some finite precision. Then we measure  $n$  times. This gives us "classes" of trajectories. We define the diameter of such a class

as  $e^{-n}$ . Now we ask: how many “classes” of trajectories can we distinguish? The scaling of the number of different classes versus its diameter defines a dimension-like quantity based on complexity arguments. This quantity is the entropy.

In order to make the situation simpler we assume that  $\mathcal{R}$  is a finite generating partition.

For every set  $Z \subset M$ , and every real number  $\delta$ , we set

$$(2) \quad N(Z, \delta, \mathcal{R}) = \liminf_{k \rightarrow \infty} \sum_{\mathcal{U} \in \Gamma_k} \exp(-\delta m(\mathcal{U})),$$

where the infimum is taken over all finite or countable collections

$$\Gamma_k \subset \bigcup_{j \geq k} \bigvee_{n=0}^{j-1} f^{-n}\mathcal{R}$$

that cover  $Z$ , and  $m(\mathcal{U}) = j$  if  $\mathcal{U} \in \bigvee_{n=0}^{j-1} f^{-n}\mathcal{R}$ .

When  $\delta$  goes from  $-\infty$  to  $+\infty$ , the quantity in (2) jumps from  $+\infty$  to 0 at a unique critical value. The number

$$h(f|Z) = \inf\{\delta : N(Z, \delta, \mathcal{R}) = 0\} = \sup\{\delta : N(Z, \delta, \mathcal{R}) = +\infty\}$$

is called the *topological entropy* of  $Z$ .

If the set  $Z$  is compact and invariant then the above definitions of entropy coincide with the classically defined topological entropies (see [10]).

The entropy with respect to the partition  $\mathcal{R}$  is

$$h(f, \mathcal{R}) := \lim_{n \rightarrow \infty} \frac{1}{n} H_\mu \left( \bigvee_{n=0}^{j-1} f^{-n}\mathcal{R} \right), \quad \text{where} \quad H_\mu(\mathcal{P}) := - \sum_{\xi \in \mathcal{P}} \mu(\xi) \log \mu(\xi).$$

The *entropy of the measure  $\mu$*  is defined as

$$h_\mu := \sup_{\mathcal{R}} h(f, \mathcal{R}).$$

This supremum is attained for generating partitions  $\mathcal{R}$ , i.e.  $h_\mu = h(f, \mathcal{R})$  and moreover  $n^{-1} H_\mu(\bigvee_{n=0}^{j-1} f^{-n}\mathcal{R})$  decreases to  $h(f, \mathcal{R})$  (see [14]). Hence

$$h_\mu = \inf_n \frac{1}{n} H_\mu \left( \bigvee_{n=0}^{j-1} f^{-n}\mathcal{R} \right)$$

for any generating partition  $\mathcal{R}$ . We can also define the *entropy of a measure  $\mu$*  as

$$h_\mu := \inf\{h(f|Z) : \mu(Z) = 1\}.$$

For *ergodic* measures it coincides with the standard definition, but it is not an affine function on the space of measures!

The very important *theorem of Shannon–McMillan–Breiman* asserts that we can compute the entropy locally for an invariant ergodic measure:

$$h_\mu = \lim_{n \rightarrow \infty} -\frac{\log \mu(C_n(x))}{n} \quad \mu\text{-a.e.},$$

where  $C_n(x)$  is the cylinder of length  $n$  containing  $x$ , that is,  $x \in C_n \in \bigvee_{k=0}^{n-1} f^{-k}\mathcal{R}$ . Note that the limit does not have to exist for all points. However, for the dimension estimates we just need the lower limit

$$(3) \quad h_\mu(x) = \liminf_{n \rightarrow \infty} -\frac{\log \mu(C_n(x))}{n}.$$

REMARK 1. The Shannon–McMillan–Breiman theorem has a very nice and useful interpretation. Let us consider the generic points for an ergodic measure  $\mu$ , i.e. all points  $x$  for which

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \phi(f^k x) = \int_M \phi d\mu$$

for all continuous functions  $\phi$ . Since  $M$  is compact, by separability and Birkhoff’s theorem this set has full measure. So its entropy is at least  $h_\mu$  (in fact, its entropy equals  $h_\mu$  by the variational principle for non-compact sets; see [10]). If  $n$  is sufficiently large then

$$\mu(C_n(x)) \sim e^{-h_\mu n},$$

and since the total mass of the cylinders covering the generic points is 1, we find that the number of “generic” (for  $\mu$ )  $n$ -cylinders is at least (approximately)  $e^{h_\mu n}$ .

If  $Z$  is compact and  $f$ -invariant, then  $h(f|Z) = h(f^{-1}|Z)$ . This is not true for non-invariant sets!

**2.3. Linear expanding maps.** We first consider the case when  $f(x) = 2x \pmod{1}$ . From now on we will drop the notation  $\pmod{1}$  since this will cause no confusion.

Multiplication by 2 yields a natural Markov partition

$$\mathbb{S}^1 = \mathbb{R}/\mathbb{Z} = I_0 \cup I_1,$$

where  $I_0 = [0, 1/2)$  and  $I_1 = [1/2, 1)$ . We can code a point  $x \in \mathbb{S}^1$  by its orbit

$$x \leftrightarrow \underline{x} = x_0 x_1 \dots x_n \dots,$$

where

$$x_n = \begin{cases} 0 & \text{if } 2^n x \in I_0, \\ 1 & \text{if } 2^n x \in I_1. \end{cases}$$

This coding corresponds to the binary coding of a real number in the unit interval. It is one-to-one except for a countable set where it is two-to-one.

Since the measures we are going to consider are non-atomic we can neglect this ambiguity and work in the coding space  $\Sigma_2 = \{0, 1\}^{\mathbb{N}}$  equipped with the product topology. We denote the coding  $\varrho: \Sigma_2 \rightarrow \mathbb{S}^1$  by  $\varrho(\underline{x}) = x$ . We define the metric

$$d^*(\underline{x}, \underline{y}) := \inf\{1/2^k : x_0 = y_0, \dots, x_k = y_k\}.$$

Pushing this metric forward to the circle we obtain a metric equivalent to the standard metric (if we neglect the points with non-unique coding). Hence we can work in the coding space endowed with the metric  $d^*$ . Moreover, the shift operator

$$\sigma(x_0x_1\dots x_n\dots) = x_1\dots x_n\dots$$

commutes with the map  $f(x) = 2x$ , i.e.  $f \circ \varrho = \varrho \circ \sigma$ .

A crucial role will be played by cylinder sets

$$C_k(\underline{x}) = \{\underline{y} \in \Sigma_2 : y_0 = x_0, \dots, y_k = x_k\}.$$

Then such a cylinder has diameter  $1/2^{k+1}$ .

We can define a metric  $d_f$  on  $\mathbb{S}^1$  as the unique metric with

$$\text{diam } I_n(\underline{y}) = \text{diam } C_n(\underline{y}).$$

**2.4. Measures on  $\mathbb{S}^1$ .** Let  $\phi: \mathbb{S}^1 \rightarrow \mathbb{R}$  be Hölder continuous. We may assume that  $P(\phi) = 0$ . The corresponding Gibbs state is the unique measure  $\mu_\phi$  with

$$C^{-1} \leq \frac{\mu_\phi(C_n(\underline{y}))}{\exp[\sum_{i=0}^n \phi(\sigma^i \underline{y})]} \leq C$$

for some  $C > 0$  and any  $n$ . Gibbs states are invariant and ergodic.

Let  $\phi$  depend only on the first coordinate:  $\phi(\underline{x}) = \phi(x_0)$ , and set

$$\phi(i) =: \phi_i, \quad e^{\phi_i} =: p_i.$$

Then  $P(\phi) = 0$  implies  $p_1 + p_2 = 1$ . The corresponding Gibbs state is called a *Bernoulli measure*.

**REMARK 2.** Gibbs states with respect to Hölder continuous potentials have approximately similar computation rules as Bernoulli measures. In the following section we will restrict our arguments to Bernoulli measures and Hölder functions which depend only on the first coordinate. In particular, we assume that the expansion rate of  $f$  is constant on the intervals  $I_1$  and  $I_2$ —i.e. the map is piecewise linear.

All results are valid for Gibbs states with arbitrary Hölder potential and for maps which are  $C^{1+\alpha}$  on  $I_i$ . Also the results can be generalised to higher-dimensional conformal systems.

**2.5. Hitting times.** Let  $f(x) = 2x \pmod{1}$  and let  $\mu_\phi$  be a Gibbs state with respect to a Hölder potential  $\phi$ . Put  $\ell_N = 1/N^\nu$ . The associated cylinder of length  $\ell_N$  in base 2 is  $C_n(y)$  with  $n \sim \nu \log N$ . We want to study the first hit time when iterates of  $x$  first hit  $C_n(y)$ . The *first hit time* of the cylinder  $C_n(y)$  is

$$\tau_n(x, y) := \min\{k : f^k(x) \in C_n(y)\}.$$

Since

$$f^n(x) \in C_n(y) \Leftrightarrow \sigma^n(x) \in C_{\nu \log N}(y) = C_n(x),$$

we need

$$\tau_n(x, y) \leq N \sim 2^{n/\nu} \quad \text{or} \quad \frac{\log \tau_n(x, y)}{n} \leq \frac{1}{\nu}$$

to have  $f^n(x)$  hitting  $C_n(y)$  infinitely often. We want to study  $\tau_n(x, y)$ . The main problem is that there is no “nice” asymptotics in general for hitting times. The case of Gibbs states is much better.

**THEOREM 1** (Chazottes, [3]). *Let  $\mu$  be an ergodic measure and  $\mu_\phi$  a Gibbs measure. For  $\mu_\phi \times \mu$  a.e.  $(x, y)$ ,*

$$\lim_{n \rightarrow \infty} \frac{\log \tau_n(x, y)}{n} = h_\mu(y) = \lim_{n \rightarrow \infty} -\frac{\log \mu_\phi(C_n(y))}{n}.$$

### 3. Standard Diophantine approximation and Diophantine classes

**DEFINITION 5.** An irrational number  $\alpha$  is of *Diophantine class*  $\nu = \nu(\alpha) \in \mathbb{R}^+$  if

$$(4) \quad \|q\alpha\| < 1/q^\mu$$

has infinitely many solutions in integers  $q$  for  $\mu < \nu$  and at most finitely many for  $\mu > \nu$ .

If  $\alpha$  is not of any Diophantine class  $\nu \in \mathbb{R}^+$  then  $\alpha$  is said to be a *Liouville number*.

If  $\alpha$  is of Diophantine class  $\nu$ , we write  $\alpha \in \text{Dioph}(\nu)$ . In a slight abuse of terminology, we say that a Liouville number  $\alpha \in (0, 1)$  has infinite Diophantine class and write  $\alpha \in \text{Dioph}(\infty)$ .

**4. Approximation by dyadic numbers.** The original motivation of this question was to find the exact speed of approximation of arbitrary  $\beta$ -shifts by subshifts of finite type. We will deal with a simpler case first.

By using the correspondence between the shift space and the circle with multiplication by 2 we address the question in the symbolic space. If no confusion can arise we will omit the correspondence and identify numbers on the circle with their dyadic expansions. Since only dyadic numbers have

two different expansions ( $01^\infty = 10^\infty$ ) we can ignore this ambiguity. We will use the metric

$$d(\underline{x}, \underline{y}) := \left| \sum_{n=1}^{\infty} \frac{x_n - y_n}{2^n} \right|.$$

This metric is equivalent to the metric on the circle. We consider a point  $\underline{x} = x_1x_2\dots \in \Sigma_2$  and ask about the speed of convergence

$$\underline{x}(n)^- := x_1\dots x_n0^\infty \rightarrow x_1x_2\dots \quad \text{or} \quad \underline{x}(n)^+ := x_1\dots x_n1^\infty \rightarrow x_1x_2\dots$$

This means we have to study the occurrence of blocks of 0's or 1's in the symbolic sequence  $\underline{x}$ . The faster such blocks occur, the better the approximation will be. Since the numbers  $x_1\dots x_n0^\infty$  and  $x_1\dots x_n1^\infty$  correspond to dyadic numbers with denominator  $2^n$  we deal with approximations of numbers by dyadic numbers. The situation here is different from the usual Diophantine approximation.

**4.1. Badly approximable numbers.** The numbers

$$\frac{j}{2^n} \pmod{1}, \quad j = 0, \dots, 2^n - 1,$$

are equally spaced on  $\mathbb{S}^1$  with distance  $1/2^n$ . Therefore for all  $x \in \mathbb{R}$  and all  $n \in \mathbb{N}$ ,  $\min_j |x - j/2^n| \leq 1/2^{n+1}$ , which gives an upper bound on the dyadic approximation.

The worst approximable numbers are

$$\underline{x}_1 = \frac{1}{3} = 01010101\dots, \quad \underline{x}_2 = \frac{2}{3} = 101010101\dots$$

They satisfy

$$r_n(x_i) := \min\{d(\underline{x}_i(n)^-, x_i), d(\underline{x}_i(n)^+, x_i)\} = \sum_{k=1}^{\infty} \frac{1}{2^{n+2k}} = \frac{1}{3 \cdot 2^n}.$$

In these two numbers neither a block of two consecutive zeros nor of two consecutive ones occurs. On the other hand, if a number  $\underline{x}$  contains infinitely many blocks of  $N > 1$  consecutive zeros or ones then

$$\min_j \left| x - \frac{j}{2^n} \right| = \min\{d(\underline{x}_i(n)^-, x_i), d(\underline{x}_i(n)^+, x_i)\} \leq \frac{1}{2^{n+N}} < \frac{1}{3 \cdot 2^n}$$

infinitely often (with equality if  $\underline{x} = x_1\dots x_{n-1}10^N1^\infty$ ). These are examples of numbers where the speed of convergence is not faster than  $2^{-n}$ . A result of Fan and Schmeling [5] yields

**THEOREM 2.**

$$\dim_{\mathbb{H}}\{x : \exists c > 0 \text{ such that } r_n(x) > c2^{-n}, \forall n \in \mathbb{N}\} = 1.$$

*Proof.* Let

$$\text{BAN} := \{x : \exists c > 0 \text{ such that } r_n(x) > c2^{-n}, \forall n \in \mathbb{N}\}$$



be the set of badly approximable numbers. We start by remarking that those numbers correspond to numbers which have bounded strings of 0's and 1's. Let  $x$  be such that there are at most  $N$  consecutive zeros or ones in its dyadic expansion. Then

$$r_n(x) = \min \left\{ \sum_{k=1}^{\infty} \frac{x_{n+k}}{2^{n+k}}, \sum_{k=1}^{\infty} \frac{1-x_{n+k}}{2^{n+k}} \right\} \geq \frac{1}{2^{n+N+1}},$$

where we have used the fact that the remainder is minimised by having a block of  $N$  consecutive zeros (or ones) starting at place  $n + 1$ , i.e.  $\underline{x} = x_1 \dots x_{n-1} 10^N 1 \dots$  or  $\underline{x} = x_1 \dots x_{n-1} 01^N 0 \dots$ . This gives  $c = 2^{N+1}$ .

Clearly, if  $N$  is unbounded, i.e. we have arbitrarily long blocks of zeros or ones, we cannot find such a  $c > 0$ .

Now the set

$$\{x : \exists c > 0 \text{ such that } r_n(x) > c2^{-n}, \forall n \in \mathbb{N}\} = \bigcup_{n \in \mathbb{N}} S_N,$$

where  $S_N$  is the subshift of finite type where blocks of  $N + 1$  consecutive zeros and ones are forbidden. If  $N \rightarrow \infty$  the dimension of  $S_N$  clearly tends to 1. ■

REMARK 3. One can “normalise” the approximation by considering the distances  $\|2^n x\|$  on the circle. By Dirichlet’s theorem, for standard continued fraction approximations and any irrational  $x \in (0, 1)$  we have

$$\min_{1 \leq p \leq q} \|px\| < 1/q, \quad \text{and hence} \quad \liminf_{q \in \mathbb{N}} \|qx\| = 0.$$

The above numbers show that this is not the case for dyadic approximations. Actually the numbers in Theorem 2 are those where

$$\liminf_{n \in \mathbb{N}} \|2^n x\| = c > 0.$$

This fact can be explained by noting that irrational rotations are uniquely ergodic and the invariant measure is Lebesgue measure independent of the irrational number  $x$ . This means that the orbit  $(qx \pmod{1})_q$  is uniformly distributed in  $\mathbb{S}^1$ . This is no longer true for multiplication by 2: then we have uncountably many different ergodic measures with completely different asymptotics.

**4.2. The “Liouville” case.** In this section we consider those numbers which have an arbitrary speed of approximation. In standard Diophantine approximation those numbers are called Liouville numbers. They form a residual subset of the real line no matter what the speed of approximation is. We are going to show that a similar statement holds for dyadic approximations.

**THEOREM 3.** *Let  $\Phi(n) \rightarrow 0$  be an arbitrary function. Then the set*

$$L_\Phi = \left\{ x \in \mathbb{R} : r_n(x) < \Phi(n) \frac{1}{2^n} \text{ infinitely often} \right\}$$

*is residual in  $\mathbb{R}$ .*

*Proof.* The proof is similar to the one for standard Diophantine approximations and is included for completeness. We note that  $\underline{x} \in L_\Phi$  if and only if there is a subsequence  $n_k$  of the naturals such that

$$x_{n_k+1} \dots x_{- \log \Phi(n_k)} = 0^{-\log \Phi(n_k)} \quad \text{or} \quad x_{n_k+1} \dots x_{- \log \Phi(n_k)} = 1^{-\log \Phi(n_k)}.$$

Clearly for each  $n \in \mathbb{N}$  the sets

$$O_n = \bigcup_{(x_1, \dots, x_n) \in \{0,1\}^n} [x_1 \dots x_n 0^{\log \Phi(n)}] \cup \bigcup_{(x_1, \dots, x_n) \in \{0,1\}^n} [x_1 \dots x_n 1^{\log \Phi(n)}]$$

are open. Moreover, the sets  $G_N = \bigcup_{n \geq N} O_n$  are dense in  $\mathbb{R}$  for all  $N \in \mathbb{N}$ . Hence, the set  $R = \bigcap_{N \in \mathbb{N}} G_N = L_\Phi$  is a dense  $G_\delta$  set, and hence is residual. ■

**REMARK 4.** As can be seen from the proof, the set of points having arbitrarily good approximations from the left (respectively from the right), i.e. with arbitrarily long sequences of zeros (respectively ones), also forms a residual set.

Let  $\Sigma_A$  be a subshift of finite type. If we code  $\Sigma_2$  with words of length  $n$  instead of just the digits 0 and 1 (which are words of length one), then if  $n$  is sufficiently large we can represent  $\Sigma_A$  by a transition matrix  $A$  of size not larger than  $2^n \times 2^n$ .

The subshift  $\Sigma_A$  is a closed subset of  $\Sigma_2 \sim [0, 1]$  and hence a Baire space. A slight modification of the above proof (just choose the cylinders  $[x_1 \dots x_n 0^{\log \Phi(n)}]$  and  $[x_1 \dots x_n 1^{\log \Phi(n)}]$  in  $\Sigma_A$ ) gives:

**THEOREM 4.** *If  $\Sigma_A$  is a subshift of finite type that is not contained in  $\{x : \exists c > 0 \text{ such that } r_n(x) > c2^{-n}, \forall n \in \mathbb{N}\}$ , i.e. arbitrarily long blocks of zeros or ones are allowed for points in  $\Sigma_A$ , then  $L_\Phi \cap \Sigma_A$  is residual in  $\Sigma_A$  and hence uncountable.*

**4.3. The “almost Liouville” case.** In this section we want to investigate those numbers which have a dyadic approximation of the order of a power of the denominator. This corresponds to the case of Diophantine numbers in the standard continued fraction approximation.

A number  $x \in \mathbb{R}$  is called *Diophantine* if there is a number  $\beta > 1$  such that

$$\|qx\| < 1/q^\beta$$

has infinitely many solutions in integers  $q$ . Jarník's theorem [7] states that for  $\beta > 1$ ,

$$\dim_{\mathbb{H}} \left\{ x \in \mathbb{R} : \|qx\| < \frac{1}{q^\beta} \text{ has infinitely many solutions } q \in \mathbb{N} \right\} = \frac{2}{\beta + 1},$$

or equivalently for  $\gamma > 2$ ,

$$\dim_{\mathbb{H}} \left\{ x \in \mathbb{R} : \left\| x - \frac{p}{q} \right\| < \frac{1}{q^\gamma} \text{ has infinitely many solutions } \frac{p}{q} \in \mathbb{Q} \right\} = \frac{2}{\gamma}.$$

Therefore for any  $\beta > 1$  or  $\gamma > 2$  the set

$$\left\{ x \in \mathbb{R} : \|qx\| < \frac{1}{q^\beta} \text{ has only finitely many solutions } q \in \mathbb{N} \right\}$$

or equivalently

$$\left\{ x \in \mathbb{R} : \left\| x - \frac{p}{q} \right\| < \frac{1}{q^\gamma} \text{ has only finitely many solutions } \frac{p}{q} \in \mathbb{Q} \right\}$$

has full Lebesgue measure.

Let  $\Sigma_A$  be a mixing subshift of finite type. We define

$$\begin{aligned} L_\alpha^A &:= \{ \underline{x} \in \Sigma_A : r_n(\underline{x}) < 2^{-\alpha n} \text{ infinitely often} \} \\ &= \left\{ x \in [0, 1] : \left| x - \frac{p}{2^n} \right| < \frac{1}{2^{\alpha n}} \right\}. \end{aligned}$$

In contrast, for dyadic approximations we have

**THEOREM 5.** *Let  $\alpha > 1$  and  $\Sigma_A$  be a mixing subshift of finite type not contained in BAN. Then*

$$\dim_{\mathbb{H}} \{ \underline{x} \in \Sigma_A : r_n(\underline{x}) < 2^{-\alpha n} \text{ infinitely often} \} = \frac{\dim_{\mathbb{H}} \Sigma_A}{\alpha}.$$

**REMARK 5.** We emphasise that in contrast to the standard Diophantine approximation we loose the 2 in the numerator, i.e. we get only half of the dimension.

*Proof.* We note first that since  $(\Sigma_2 \setminus \text{BAN}) \cap \Sigma_A \neq \emptyset$ , arbitrarily long sequences of zeros or ones are allowed in  $\Sigma_A$ . Let  $T > 0$  be such that  $A^T > 0$ . A number  $\underline{x}$  is in the set  $L_\alpha^A$  if and only if there is a sequence  $n_k(\underline{x})$  such that  $r_{n_k}(\underline{x}) < 2^{-\alpha n_k}$ , i.e.  $x_{n_k+1} = 0, \dots, x_{n_k+(\alpha-1)n_k} = 0$  (or ones respectively). Let us fix a sequence  $(n_k)_k$  such that

$$\sum_{i=0}^{k-1} (2T + (\alpha - 1)n_i) < \log n_k.$$

We consider the following map  $g$  from

$$L_\alpha^A(n_k) := \{ \underline{x} \in \Sigma_A : r_{n_k}(\underline{x}) < 2^{-\alpha n_k}, \forall k \in \mathbb{N} \}$$

into  $\Sigma_2$ . For  $\underline{x} \in L_\alpha^A(n_k)$  we will collapse the sequence of zeros (or ones) occurring at the places  $n_k$ . More precisely, the point

$$\underline{x} = x_1 \dots x_{n_k-T-1} x_{n_k-T} \dots x_{n_k} 0^{(\alpha-1)n_k} x_{n_k+(\alpha-1)n_k+1} \dots \\ \dots x_{n_k+(\alpha-1)n_k+T-1} x_{n_k+(\alpha-1)n_k+T} \dots$$

is sent by  $g$  to

$$g(\underline{x}) = x_1 \dots x_{n_k-T} x_{n_k+(\alpha-1)n_k+T+1} \dots$$

This point does not have to belong to  $\Sigma_A$  since transition from  $x_1 \dots x_{n_k-T-1}$  to  $x_{n_k+(\alpha-1)n_k+T} \dots$  might be forbidden in  $\Sigma_A$ . However, if

$$\underline{y} = y_1 \dots y_{n_k-T-1} \dots \in \Sigma_A$$

then

$$\underline{x} = y_1 \dots y_{n_k} z_1 \dots z_T 0^{(\alpha-1)n_k} w_1 \dots w_T y_{n_k+1} \dots \in L_\alpha^A(n_k)$$

for suitably chosen  $z_1 \dots z_T$  and  $w_1 \dots w_T$ . Hence,

$$\Sigma_A \subset g(L_\alpha^A(n_k)).$$

We are going to estimate the Hölder exponent of  $g$ . For simplicity we will use the metric  $d^*$  which has equivalent diameters of cylinder sets. Let  $\underline{x}, \underline{y} \in L_\alpha^A(n_k)$ . Then their distance  $d^*(\underline{x}, \underline{y})$  is determined by the first digit where these sequences differ; say  $x_m \neq y_m$  ( $x_i = y_i$  for  $1 \leq i < m$ ) and let  $k$  be such that  $\alpha n_k \leq m < n_{k+1}$  ( $m$  cannot be in the intervals  $(n_k, \alpha n_k)$  since there both sequences are 0). Then for  $g(\underline{x})$  and  $g(\underline{y})$  we erase at most  $\sum_{i=0}^{k-1} (2T + (\alpha - 1)n_i) < \log n_k$  symbols before  $n_k - T < m$  and at most  $2T + (\alpha - 1)n_k$  symbols between  $n_k - T$  and  $m$ . Thus for some  $m' > m - (\log n_k + 2T + (\alpha - 1)n_k)$  we have  $g(\underline{x})_i = g(\underline{y})_i$  for  $1 \leq i \leq m'$  and  $g(\underline{x})_{m'} \neq g(\underline{y})_{m'}$ . This yields, for sufficiently large  $k$ ,

$$d^*(g(\underline{x}), g(\underline{y})) = 2^{-m'} \leq 2^{-m+(\log n_k+2T+(\alpha-1)n_k)} \leq 2^{-m+(\alpha-1+\varepsilon)n_k} \\ \leq 2^{-m+(\alpha-1+\varepsilon)m/\alpha} \leq (2^{-m})^{(1-\varepsilon)/\alpha}.$$

Since a Hölder continuous map with Hölder exponent  $\kappa$  can increase the dimension at most by a factor of  $1/\kappa$ , we get

$$\dim_H \Sigma_A \leq \frac{\alpha}{1-\varepsilon} \dim_H L_\alpha^A(n_k) \leq \frac{\alpha}{1-\varepsilon} \dim_H L_\alpha^A.$$

Letting  $\varepsilon \rightarrow 0$  gives the right lower bound.

For the upper bound we remark that  $L_\alpha^A = \bigcup_{(n_k)_k} L_\alpha^A(n_k)$  and  $L_\alpha^A(n_k)$  can be covered by cylinders of length  $\alpha n_k$  for each  $k \in \mathbb{N}$ . The number of those cylinders is at most the number of cylinders of length  $n_k$  intersecting  $\Sigma_A$ . But for each  $\varepsilon > 0$  and  $k$  sufficiently large this number is bounded by

$2^{(\dim_{\mathbb{H}} \Sigma_A + \varepsilon) \cdot n_k}$ . Hence, for those covering sums we get

$$\begin{aligned} \sum_{\substack{C_i \cap L_\alpha^A \neq \emptyset \\ C_i \cap C_j = \emptyset}} \text{diam } C_i^{(\dim_{\mathbb{H}} \Sigma_A)/\alpha + \varepsilon} &\leq \sum_{n \in \mathbb{N}} \sum_{\substack{C_i \cap L_\alpha^A \neq \emptyset \\ |C_i|=n}} \text{diam } C_i^{(\dim_{\mathbb{H}} \Sigma_A)/\alpha + \varepsilon} \\ &\leq \sum_{n \in \mathbb{N}} 2^{(\dim_{\mathbb{H}} \Sigma_A + \varepsilon) \cdot n} \left( \frac{1}{2^{\alpha n}} \right)^{(\dim_{\mathbb{H}} \Sigma_A)/\alpha + \varepsilon} \\ &\leq \sum_{n \in \mathbb{N}} 2^{(1-\alpha)\varepsilon n} < \infty. \end{aligned}$$

This shows that  $\mathcal{H}^s(L_\alpha^A) < \infty$ , where  $s = (\dim_{\mathbb{H}} \Sigma_A)/\alpha + \varepsilon$ . (Actually we estimated the upper box-counting measure). By Definition 3 it follows that  $\dim_{\mathbb{H}} L_\alpha^A \leq (\dim_{\mathbb{H}} \Sigma_A)/\alpha + \varepsilon$ . Since  $\varepsilon > 0$  is arbitrary we conclude that  $\dim_{\mathbb{H}} L_\alpha^A \leq (\dim_{\mathbb{H}} \Sigma_A)/\alpha$ . ■

REMARK 6. Those numbers have the property that infinitely often

$$x_n = 0, \dots, x_{\alpha n} = 0 \quad \text{or} \quad x_n = 1, \dots, x_{\alpha n} = 1.$$

Therefore no invariant measure besides the delta measure on the sequence  $0^\infty$ , respectively  $1^\infty$ , can sit on this set. To see this, we remark that by Birkhoff’s ergodic theorem the frequencies

$$F(\underline{x}, i) := \lim_{N \rightarrow \infty} \frac{\#\{1 \leq n \leq N : x_n = i\}}{N}$$

for  $i = 0, 1$  have to exist for an invariant measure  $\mu$  and for  $\mu$ -a.e.  $\underline{x}$ . We claim that the frequency of 0’s (or 1’s) must be 1.

Assume that  $\mu$  is an invariant measure on the set of points with

$$x_n = 0, \dots, x_{\alpha n} = 0$$

infinitely often. Then for  $\mu$ -a.e.  $\underline{x}$ ,  $F(\underline{x}, 0)$  exists. If  $n_k$  is an infinite sequence such that  $x_{n_k} \dots x_{\alpha n_k} = 0$  then

$$F_{\alpha n_k}(\underline{x}, 0) = \frac{n_k F_{n_k}(\underline{x}, 0) + (\alpha - 1)n_k}{\alpha n_k},$$

where  $F(\underline{x}, i)_N := \#\{1 \leq n \leq N : x_n = i\}/N$ . Hence we have

$$\alpha F_{\alpha n_k}(\underline{x}, 0) - F_{\alpha n_k}(\underline{x}, 0) = \alpha - 1.$$

Since  $F_N(\underline{x}, 0)$  converges to  $F(\underline{x}, 0)$  we must have  $F(\underline{x}, 0) = 1$ . Now  $\delta_{0^\infty}$  is the only invariant measure having frequency one of the digit 0.

A similar argument holds for long blocks of ones.

**4.4. The “Diophantine case”.** By Chazotte’s theorem (Theorem 1), we know that for equilibrium measures  $\mu_\phi$  the hitting times have the following asymptotics (in logarithmic scale):

$$\tau_k(\underline{x}, 0^\infty) \sim 2^{h_{\mu_\phi}(0^\infty)k} \quad \text{and} \quad \tau_k(\underline{x}, 1^\infty) \sim 2^{h_{\mu_\phi}(1^\infty)k}$$

for  $\mu_\phi$ -a.e.  $\underline{x}$ , where  $h_{\mu_\phi}(0^\infty)$  and  $h_{\mu_\phi}(1^\infty)$  are defined by (3). So we expect, for a point  $\underline{x}$  typical with respect to  $\mu_\phi$ , to have a block of  $k$  consecutive zeros (or ones) at place approximately  $2^{h_{\mu_\phi}(0^\infty)k}$ , or equivalently a block of  $h_{\mu_\phi}^{-1} \log n$  consecutive zeros at  $n = 2^{h_{\mu_\phi}(0^\infty)k}$ . This means that we should expect (with  $0 < \beta = 1/\min\{h_{\mu_\phi}(0^\infty), h_{\mu_\phi}(1^\infty)\} < 1$ )

$$r_n(x) < \frac{1}{2^{n\beta}} \quad \text{infinitely often.}$$

Clearly according to the previous theorem these sets have full dimension.

For  $\beta > 0$  we are interested when this approximation speed is the best possible. This is, we are interested in the set

$$\begin{aligned} D_\beta &:= \left\{ \underline{x} \in \Sigma_2 : r_n(\underline{x}) < \frac{1}{2^{n\beta+\varepsilon}} \text{ finitely often for any } \varepsilon > 0 \text{ and } \right. \\ &\quad \left. r_n(\underline{x}) < \frac{1}{2^{n\beta-\varepsilon}} \text{ infinitely often for any } \varepsilon < 0 \right\} \\ &= \left\{ \underline{x} : \limsup_{n \rightarrow \infty} -\frac{\log 2^n r_n(\underline{x})}{\log n} = \beta \right\}. \end{aligned}$$

This corresponds to the Diophantine classes for standard Diophantine approximation (see Sections 3 and 4.3).

For dyadic approximations we have the following

**THEOREM 6.** *For any  $0 < \beta < 1$ ,*

$$\dim_{\mathbb{H}} \left\{ \underline{x} : \limsup_{n \rightarrow \infty} -\frac{\log 2^n r_n(\underline{x})}{\log n} = \beta \right\} = 1.$$

*Proof.* This proof follows the methods developed in [5, Theorem 4.1]. Let us consider the set

$$\text{BAN} = \{ \underline{x} \in \Sigma_2 : \exists c > 0 \text{ such that } r_n(\underline{x}) > c2^{-n}, \forall n \in \mathbb{N} \}.$$

Then by Theorem 2 this set has full dimension. Also this set consists of sequences  $\underline{x}$  having bounded sequences of both zeros and ones.

Now let  $n_k = 2^{k/\beta}$ . We are going to modify the sequences in BAN. Let  $\underline{x} \in \text{BAN}$ . For  $k \in \mathbb{N}$  we insert a block of  $k$  consecutive zeros between  $x_{n_k}$  and  $x_{n_{k+1}}$ , i.e.

$$x_1 \dots x_{n_k} x_{n_k+1} \dots \mapsto x_1 \dots x_{n_k} 0^k x_{n_k+1} \dots$$

By the proof of Theorem 4.1 in [5] ( $k \ll n_k$ ) this procedure does not decrease the dimension. Therefore the resulting set  $\text{BAN}_{\text{modified}}$  has dimension 1. Clearly the fastest approximation is obtained exactly at the places  $n_k$ . Hence for  $\underline{x}_{\text{modified}} \in \text{BAN}_{\text{modified}}$  we have  $\limsup_n -\frac{\log 2^n r_n(\underline{x})}{\log n} = \beta$ . ■

**REMARK 7.** This means that we have uncountably many disjoint sets of full dimension with different approximation speed.

**4.5.** *The “regular Diophantine case”.* The dynamical system underlying the standard Diophantine approximation is the rotation. For an irrational number  $x$ , rotation by angle  $x: x \mapsto nx \pmod{1}$  has a unique invariant measure, the Lebesgue measure on  $\mathbb{S}^1$ . By Birkhoff’s ergodic theorem

$$\lim_{n \rightarrow \infty} \frac{1}{n} \# \left\{ 0 \leq k < n : \|kx\| < \frac{1}{q} \right\} = \frac{2}{q}.$$

Hence, we can “regularise” the approximation by setting

$$\tau_q(x) := \frac{\log \lim_{n \rightarrow \infty} \frac{1}{n} \# \left\{ 0 \leq k < n : \|kx\| < \frac{1}{q} \right\}}{-\log q}$$

and

$$\tau(x) := \lim_{q \rightarrow \infty} \tau_q(x).$$

This number gives the average approximation rate over the orbit of  $x$ . From the discussion above it follows that  $\tau(x) = 1$  for all irrational  $x$ . This equals the Diophantine approximation rate if the Diophantine class  $\beta(x)$  is 1. It is well known that the set of numbers for which  $\beta(x) = \tau(x) = 1$  has full Lebesgue measure. In general, we have

$$1 = \tau(x) \leq \beta(x).$$

The situation in dyadic Diophantine approximation is different. Here we define

$$\tau_m^{\text{reg}}(x) := \frac{\log \liminf_{N \rightarrow \infty} \frac{1}{N} \# \left\{ 0 \leq k < N : \|2^k x\| < \frac{1}{2^m} \right\}}{-m}$$

and

$$\tau^{\text{reg}}(x) := \inf_m \tau_m^{\text{reg}}(x).$$

In symbolic language this is

$$\tau_m^{\text{reg}}(\underline{x}) = \frac{\log \liminf_{N \rightarrow \infty} \max \{ F_N(\underline{x}, 0^m), F_N(\underline{x}, 1^m) \}}{-m}.$$

We are interested in the size of the set of points

$$D_\beta^{\text{reg}} := \{ \underline{x} \in \Sigma_2 : \tau^{\text{reg}}(\underline{x}) = \log \beta \} \quad \text{for } \beta \geq 2.$$

Let  $\underline{x} \in D_\beta^{\text{reg}}$ . Then for each  $m \in \mathbb{N}$  there is a sequence  $(n_k^{(m)})_k$  such that

$$\frac{\log \lim_{k \rightarrow \infty} \max \{ F_{n_k^{(m)}}(\underline{x}, 0^m), F_{n_k^{(m)}}(\underline{x}, 1^m) \}}{-m} = \log \beta_m$$

and  $\inf_m \beta_m = \beta$ . By using a diagonal argument we can choose a universal sequence  $(n_k)_k$  along which all the limits exist. Consider the set

$$V_\sigma(\underline{x}) := \left\{ \mu \text{ inv.} : \exists (n_k)_k \text{ such that } \lim_{k \rightarrow \infty} \frac{1}{n_k} \sum_{i=0}^{n_k-1} \delta_{\sigma^i \underline{x}} = \mu \right\}.$$

By using subsequences of the universal sequence  $(n_k)_k$  from above we derive that there is a measure  $\mu \in V_\sigma(\underline{x})$  with

$$\mu([0^m]) = \beta_m^{-m} \quad \text{or} \quad \mu([1^m]) = \beta_m^{-m}.$$

Therefore,

$$\max\left\{-\frac{1}{m} \log \mu([0^m]), -\frac{1}{m} \log \mu([1^m])\right\} \geq \log \beta.$$

We are going to use the following result of Bowen [2].

**THEOREM 7 (Bowen).**

$$h_{\text{top}}\{\underline{x} \in \Sigma_2 : \exists \mu \in V_\sigma(\underline{x}) \text{ with } h_\mu \leq t\} \leq t.$$

As observed above, if  $\underline{x} \in D_\beta^{\text{reg}}$  then there exists a  $\mu \in V_\sigma(\underline{x})$  such that  $\max\{-\frac{1}{m} \log \mu([0^m]), -\frac{1}{m} \log \mu([1^m])\} \geq \log \beta$  for all  $m$ . This shows that

$$\begin{aligned} & \sup_{\underline{x} \in D_\beta^{\text{reg}}} \{ \inf \{ h_\mu : \mu \in V_\sigma(\underline{x}) \} \} \\ & \leq S := \sup_{\underline{x} \in D_\beta^{\text{reg}}} \left\{ \inf \left\{ h_\mu : \max \left\{ -\frac{1}{m} \log \mu([0^m]), -\frac{1}{m} \log \mu([1^m]) \right\} \right. \right. \\ & \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \left. \geq \log \beta, \forall m \right\} \left. \right\}. \end{aligned}$$

Then  $D_\beta^{\text{reg}} \subset \{ \underline{x} : \exists \mu \in V_\sigma(\underline{x}) \text{ with } h_\mu \leq S \}$ . Now Theorem 7 implies

**PROPOSITION 1.** *Let  $\beta \geq 2$ . Then*

$$\begin{aligned} & \dim_{\text{H}} D_\beta^{\text{reg}} \\ & \leq \sup \left\{ h_\mu \text{ inv.} : \max \left\{ -\frac{1}{m} \log \mu([0^m]), -\frac{1}{m} \log \mu([1^m]) \right\} \right. \\ & \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \left. \geq \log \beta, \forall m \right\}. \end{aligned}$$

The next step is to evaluate the supremum over those measures.

**THEOREM 8.** *Let  $\beta \geq 2$ . Then*

$$\begin{aligned} & \sup \left\{ h_\mu \text{ inv.} : \max \left\{ -\frac{1}{m} \log \mu([0^m]), -\frac{1}{m} \log \mu([1^m]) \right\} \right. \\ & \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \left. \geq \log \beta, \forall m \right\} \\ & = -\frac{1}{\beta} \log \frac{1}{\beta} - \left( 1 - \frac{1}{\beta} \right) \log \left( 1 - \frac{1}{\beta} \right). \end{aligned}$$

*This means that the dimension is carried by a Bernoulli measure.*



*Proof.* Clearly,

$$\begin{aligned} \sup \left\{ h_\mu \text{ inv.} : \max \left\{ -\frac{1}{m} \log \mu([0^m]), -\frac{1}{m} \log \mu([1^m]) \right\} \geq \log \beta, \forall m \right\} \\ \leq \sup \{ h_\mu \text{ inv.} : \max \{ -\log \mu([0]), -\log \mu([1]) \} \geq \log \beta \} \\ = \sup \{ h_\mu \text{ inv.} : \min \{ \mu([0]), \mu([1]) \} \leq \beta^{-1} \}. \end{aligned}$$

But  $\mu([0]) = 1 - \mu([1])$ . Without loss of generality we may assume that  $\mu([0]) \leq \beta^{-1} \leq 12$ . Then

$$h_\mu \leq H_\mu(\{[0], [1]\}) \leq -\frac{1}{\beta} \log \frac{1}{\beta} - \left(1 - \frac{1}{\beta}\right) \log \left(1 - \frac{1}{\beta}\right).$$

The lower bound is obtained by considering the  $(1/\beta, 1 - 1/\beta)$ -Bernoulli measure  $\mu_\beta$  which sits on the set in question since  $\mu_\beta([0^m]) = \beta^{-m}$  and  $\mu_\beta([1^m]) = (1 - \beta^{-1})^{-m}$ . ■

REMARK 8. In the definition of  $\tau^{\text{reg}}$  we cannot replace inf by lim inf. Consider the set BAN. There  $\liminf_{n \rightarrow \infty} \tau_m^{\text{reg}}(\underline{x}) = \infty$ , implying that  $-\frac{1}{\beta} \log \frac{1}{\beta} - (1 - \frac{1}{\beta}) \log(1 - \frac{1}{\beta}) = 0$  but the dimension of BAN is 1.

REMARK 9. We have proved that a maximal ergodic measure is sitting on the sets with regular approximation. Therefore for a.e. point with respect to this measure, we can actually replace lim inf and inf in the definitions of  $\tau_m^{\text{reg}}$  and  $\tau^{\text{reg}}$  with limits.

REMARK 10. The Lebesgue measure on the circle is obtained by pulling back the  $(1/2, 1/2)$ -Bernoulli measure. Therefore

$$\text{Leb}(D_2^{\text{reg}}) = \dim_{\mathbb{H}} \{x \in \mathbb{S}^1 : \tau^{\text{reg}}(x) = \log 2\} = 1.$$

REMARK 11. If we have an ergodic measure  $\mu$  then

$$\tau^{\text{reg}}(\underline{x}) = \max \{ h_\mu(0^\infty), h_\mu(1^\infty) \} \quad \mu\text{-a.e. } \underline{x},$$

where

$$h_\mu(\underline{y}) = \liminf_{n \rightarrow \infty} -\frac{1}{n} \log \mu(C_n(\underline{y})).$$

In [6] it was proved (as a slight modification of Chazotte’s theorem) that for an equilibrium state  $\mu$  with respect to a Hölder continuous potential,

$$\lim_{m \rightarrow \infty} \frac{1}{m} \log \min \{ n > 1 : \sigma^n(\underline{x}) \in C_m(\underline{y}) \} = h_\mu(\underline{y}) \quad \mu\text{-a.e. } \underline{x}.$$

Therefore for  $\beta \geq 2$ ,

$$\begin{aligned} \dim_{\mathbb{H}} \left\{ x \in \mathbb{S}^1 : r(n) \sim 2^{-n} 2^{-(\log n)/\beta} = \frac{1}{2^n n^{1/\beta}}, n > n_0(x) \right\} \\ = -\frac{1}{\beta} \log \frac{1}{\beta} - \left(1 - \frac{1}{\beta}\right) \log \left(1 - \frac{1}{\beta}\right). \end{aligned}$$

Moreover, we have a maximising ergodic measure sitting on these sets. In particular,

$$\text{Leb}\left\{x \in \mathbb{S}^1 : r(n) \sim \frac{1}{2^{n_n}}, n > n_0(x)\right\} = 1.$$

**5. Approximation by finite  $\beta$ -expansions.** Let  $1 < \beta < 2$  and let  $T_\beta: [0, 1] \rightarrow [0, 1]$  be the  $\beta$ -expansion defined by

$$T_\beta: x \mapsto \beta x \text{ mod } 1.$$

Denote by  $i(x, \beta) = \{i_n(x, \beta)\}_{n=0}^\infty \in \{0, 1\}^\mathbb{N}$  the sequence defined by

$$i_n(x, \beta) = [\beta T_\beta^n(x)].$$

This sequence is called the *expansion of  $x$  in base  $\beta$* . It satisfies

$$x = \sum_{n=0}^\infty \frac{i_n(x, \beta)}{\beta^{n+1}}.$$

The set  $S_\beta = \text{closure}\{i(x, \beta) : x \in [0, 1]\}$  together with the left shift  $\sigma: \{i_n\}_{n=0}^\infty \mapsto \{i_{n+1}\}_{n=0}^\infty$  is called the  *$\beta$ -shift*.  $S_\beta$  can be characterised by

$$S_\beta = \{\{j_n\}_{n=0}^\infty \in \{0, 1\}^\mathbb{N} : \sigma^k\{j_n\} < i(1, \beta), \forall k \geq 0\},$$

where  $<$  denotes the lexicographical ordering. Let  $\Xi$  be the map  $\beta \mapsto i(1, \beta)$ . A sequence  $\{i_n\}$  is an expansion of 1 in some base  $\beta$  if and only if  $\{i_n\}$  satisfies  $\sigma^k\{i_n\} < \{i_n\}$  for any  $k > 0$ . Sequences that satisfies  $\sigma^k\{i_n\} < \{i_n\}$  for any  $k > 0$  will be called *admissible*. Moreover,  $\beta_1 < \beta_2$  if and only if  $i(1, \beta_1) < i(1, \beta_2)$ .

The subshift  $S_\beta$  is of finite type if and only if the sequence  $i(1, \beta)$  is either periodic or terminates with zeros.

For more details on the  $\beta$ -expansion, see for example [1, 9, 13].

REMARK 12. We can interpret the forthcoming estimates for the Diophantine approximation in  $S_\beta$  as the “speed of convergence” of subshifts of finite type  $S_{\beta'}$  to  $S_\beta$ . Namely, we investigate the approximation of  $i(1, \beta) = i_0i_1 \dots i_n i_{n+1} \dots$  by sequences  $i(1, \beta') = i_0i_1 \dots i_{n-1}0^\infty$ . The latter corresponds to a  $S_{\beta'}$  of finite type inside  $S_\beta$ , so the approximation speed determines how closely  $S_\beta$  is approximated by subshifts of finite type with a given size of the transition matrix. (For  $\beta'$ , the transition matrix  $A$  can be chosen to be of size at most  $(n + 1) \times (n + 1)$ .) This is an important issue since the size of  $A$  is responsible for the bounding constants appearing in many statistical properties of a subshift of finite type. (Since  $S_{\beta'} \subset S_\beta$ , the exponential growth rates are bounded by those of  $S_\beta$ .)

**5.1. Extension of theorems to the  $\beta$ -expansion.** If in Section 4 we replace the full two-shift  $\Sigma_2$  with the  $\beta$ -shift we can state theorems similar to those in Section 4. Since for  $\beta < 2$  the maximal number of consecutive 1’s in  $S_\beta$

is bounded we will get

$$r_n(x) := d(\underline{x}(n)^-, \underline{x}).$$

The following theorems are analogous to those in Section 4, hence we state them without proof.

**THEOREM 9.**

$$\dim_{\mathbb{H}}\{x : \exists c > 0 \text{ such that } r_n(x) > c\beta^{-n}, \forall n \in \mathbb{N}\} = 1.$$

**THEOREM 10.** *Let  $\Phi(n) \rightarrow 0$  be an arbitrary function. Then the set*

$$L_{\Phi} = \left\{ x \in \mathbb{R} : r_n(x) < \Phi(n) \frac{1}{\beta^n} \text{ infinitely often} \right\}$$

*is residual in  $\mathbb{R}$ .*

**THEOREM 11.** *Let  $\alpha > 1$  and  $\Sigma_A \subset S_{\beta}$  be a subshift of finite type not contained in BAN. Then*

$$\dim_{\mathbb{H}}\{\underline{x} \in \Sigma_A : r_n(\underline{x}) < \beta^{-\alpha n} \text{ infinitely often}\} = \frac{\dim_{\mathbb{H}} \Sigma_A}{\alpha}.$$

**THEOREM 12.** *For any  $0 < \alpha < 1$ ,*

$$\dim_{\mathbb{H}}\left\{ \underline{x} : \limsup_n -\frac{\log \beta^n r_n(\underline{x})}{\log n} = \alpha \right\} = 1.$$

**THEOREM 13.** *Let  $\alpha \geq 2$ . Then*

$$\begin{aligned} \sup \left\{ h_{\mu} \text{ inv.} : -\frac{1}{m} \log \mu([0^m]) \geq \log \alpha \right\} \\ = -\frac{1}{\alpha} \log \frac{1}{\alpha} - \left(1 - \frac{1}{\alpha}\right) \log \left(1 - \frac{1}{\alpha}\right). \end{aligned}$$

**5.2. Diophantine approximation with  $\beta$ -expansions.** A cylinder, denoted by  $C_n([i_0 \dots i_{n-1}], \beta)$ , is the set

$$C_n([i_0 \dots i_{n-1}], \beta) = \{x \in [0, 1] : i_k(x, \beta) = i_k, k = 0, \dots, n - 1\}.$$

The collection of all cylinders of length  $n$  is denoted  $\mathcal{C}_n$ . A cylinder in the parameter space is the set

$$C_n^{\mathbb{P}}([i_0 \dots i_{n-1}]) = \{\beta \in (1, \infty) : i_k(1, \beta) = i_k, k = 0, \dots, n - 1\}.$$

The collection of all cylinders of length  $n$  in the parameter space is denoted  $\mathcal{C}_n^{\mathbb{P}}$ .

Let  $\alpha > 0$  and fix  $\beta_0 < \beta_1 \in (1, 2)$ . We are going to calculate the dimension of the set

$$\begin{aligned} D_{\alpha}(\beta_0, \beta_1) &= \{\beta \in (\beta_0, \beta_1) : d(\sigma^n \Xi(\beta), \{0\}^{\mathbb{N}}) < \beta^{-\alpha n} \text{ infinitely often}\} \\ &= \{\beta \in (\beta_0, \beta_1) : i_n(1, \beta) \dots i_{(1+\alpha)n}(1, \beta) = 0 \dots 0 \text{ infinitely often}\}. \end{aligned}$$

For this purpose we will need the following lemmata. We start by bounding the dimension of  $D_{\alpha}(\beta_0, \beta_1)$  from above.

LEMMA 2.  $\dim_{\mathbb{H}}(D_{\alpha}(\beta_0, \beta_1)) \leq 1/(1 + \alpha)$ .

*Proof.* Let  $K_{\beta_1}$  be the set of  $x \in [0, 1]$  such that  $i(x, \beta_1) = \Xi(\beta)$  for some  $\beta \in (1, \beta_1)$ . Let  $\varrho_{\beta_1}: K_{\beta_1} \rightarrow (1, 2)$  be the map  $x \mapsto \beta$ . We estimate the dimension of  $\varrho_{\beta_1}^{-1}(D_{\alpha})$  from above.

The set

$$B_{\alpha} = \bigcap_{N=1}^{\infty} \bigcup_{n \geq N} \bigcup_i [i_0 \dots i_{n-1} \underbrace{0 \dots 0}_{\alpha n}] \subset S_{\beta_1}$$

satisfies  $\varrho_{\beta_1}(K_{\beta_1} \cap B_{\alpha}) \supset D_{\alpha}(\beta_0, \beta_1)$ . It follows that  $\dim_{\mathbb{H}}(D_{\alpha}(\beta_0, \beta_1)) \leq \dim_{\mathbb{H}} \varrho_{\beta_1}(K_{\beta_1} \cap B_{\alpha})$ .

There are constants  $c_0, c_1 > 1$  such that the number of cylinders of length  $n$  in  $S_{\beta_1}$  is bounded by  $c_0 \beta_1^n$  and the length of a cylinder of length  $n$  is bounded by  $c_1 \beta_1^{-n}$ . We use this to bound the Hausdorff dimension of  $B_{\alpha}$  from above by  $1/(1 + \alpha)$ . The set  $B_{\alpha}$  is covered by the collection of all cylinders of the form  $C_{(1+\alpha)n}([i_0 \dots i_{n-1} 0 \dots 0], \beta_1)$  where  $n$  is larger than some number  $n_0$ . For any  $\varepsilon > 0$ ,

$$\begin{aligned} \liminf_{N \rightarrow \infty} \sum_{n \geq N} \sum_i |C_{(1+\alpha)n}([i_0 i_1 \dots i_{n-1} 0 \dots 0], \beta_1)|^{(1+\varepsilon)/(1+\alpha)} \\ \leq \lim_{N \rightarrow \infty} \sum_{n \geq N} c_0 \beta_1^n (c_1 \beta_1^{-(1+\alpha)n})^{(1+\varepsilon)/(1+\alpha)} = 0. \end{aligned}$$

This implies that  $\dim_{\mathbb{H}}(B_{\alpha}) \leq (1 + \varepsilon)/(1 + \alpha)$ . Hence  $\dim_{\mathbb{H}}(D_{\alpha}) \leq 1/(1 + \alpha)$ .

The estimate of the Hölder exponent of the map  $\varrho_{\beta_1}$  in [13] implies that

$$\dim_{\mathbb{H}}(\varrho_{\beta_1}(B_{\alpha})) \leq \frac{\log \beta_1}{\log \beta_0} \dim_{\mathbb{H}}(B_{\alpha}) \leq \frac{1}{1 + \alpha} \frac{\log \beta_1}{\log \beta_0}$$

and so  $\dim_{\mathbb{H}}(D_{\alpha}(\beta_0, \beta_1)) \leq \frac{1}{1 + \alpha} \frac{\log \beta_1}{\log \beta_0}$ .

Given any  $\varepsilon > 0$  we can decompose the interval  $(\beta_0, \beta_1)$  into finitely many intervals  $(\beta_0, \beta_1) = \bigcup_k (\beta_{0,k}, \beta_{1,k})$  with  $\log \beta_{1,k} / \log \beta_{0,k} < 1 + \varepsilon$ . Then

$$\dim_{\mathbb{H}}(D_{\alpha}(\beta_0, \beta_1)) = \max_k \dim_{\mathbb{H}}(D_{\alpha}(\beta_{0,k}, \beta_{1,k})) \leq \frac{1 + \varepsilon}{1 + \alpha}.$$

We conclude that  $\dim_{\mathbb{H}}(D_{\alpha}(\beta_0, \beta_1)) \leq 1/(1 + \alpha)$ . ■

It remains to bound the dimension of  $D_{\alpha}(\beta_0, \beta_1)$  from below. A difficulty in doing this is that even though most cylinders of size  $n$  in  $C_n^p$  have diameter approximately  $\beta^{-n}$ , where  $\beta$  is in the cylinder, there are some cylinders of size  $n$  that have much smaller diameter. It is important that these small cylinders do not accumulate too much. The following lemma gives us sufficient control over the small cylinders.

LEMMA 3. *Among any  $n$  consecutive cylinders in  $C_n^p$  there is at least one  $C_n^p$  with  $|C_n^p| > \beta^{-n}$ , where  $\beta = \sup C_n^p$ .*

*Proof.* Fix a natural number  $n$  and take any  $n$  consecutive cylinders in  $C_n^{\mathbb{P}}$ . Denote by  $C_n^{\mathbb{P}}([i_0 \dots i_{n-1}])$  the cylinder, among the  $n$  consecutive cylinders, containing the smallest numbers. Then the sequence  $i_0 \dots i_{n-1}00 \dots$  is admissible and it is contained in the cylinder  $C_n^{\mathbb{P}}([i_0 \dots i_{n-1}])$ . If the sequence  $i_0 \dots i_{n-1}100 \dots$  is admissible then we are done, since then, if  $i_0 \dots i_{n-1}00 \dots = i(1, \beta_0)$  and  $i_0 \dots i_{n-1}100 \dots = i(1, \beta_1)$ , we have

$$(5) \quad \begin{aligned} \beta_1 - \beta_0 &= \sum_{k=0}^{\infty} \left( \frac{i_k(1, \beta_1)}{\beta_1^k} - \frac{i_k(1, \beta_0)}{\beta_0^k} \right) \\ &\geq \sum_{k=0}^{\infty} \left( \frac{i_k(1, \beta_1)}{\beta_1^k} - \frac{i_k(1, \beta_0)}{\beta_1^k} \right) \geq \beta_1^{-n} \end{aligned}$$

and  $\beta_1$  and  $\beta_0$  are in  $C_n^{\mathbb{P}}([i_0 \dots i_{n-1}])$ . Hence the diameter of the cylinder  $C_n^{\mathbb{P}}([i_0 \dots i_{n-1}])$  is at least  $\beta_1 - \beta_0 \geq \beta_1^{-n}$ .

We now assume that the sequence  $i_0 \dots i_{n-1}100 \dots$  is not admissible. Then there must exist maximal  $k_1 < n$  such that

$$i_{k_1} \dots i_{n-1}0 = i_0 \dots i_{n-k_1},$$

as otherwise the sequence  $i_0 \dots i_{n-1}100 \dots$  would be admissible. By the maximality of  $k_1$ , the sequence

$$(i_0 \dots i_{n-k_1})(i_0 \dots i_{n-k_1}) \dots$$

is admissible and so  $i_0 \dots i_{n-k_1-1}100 \dots$  is admissible as well, since

$$i_0 \dots i_{n-k_1-1}1 = i_0 \dots i_{n-k_1}$$

and  $i_0 \dots i_{n-k_1}$  is admissible. The two cylinders

$$C_n^{\mathbb{P}}([i_0 \dots i_{n-1}]) \quad \text{and} \quad C_n^{\mathbb{P}}([i_0 \dots i_{n-k_1-1}10 \dots 0])$$

are neighbours and therefore they are among the  $n$  consecutive cylinders.

We will now argue for the sequence  $i_0 \dots i_{n-k_1-1}100 \dots$  just as above for  $i_0 \dots i_{n-1}00 \dots$ . The sequence  $i_0 \dots i_{n-k_1-1}100 \dots$  is admissible, and if the sequence

$$i_0 \dots i_{n-k_1-1}100 \dots 0100 \dots,$$

where the last 1 is in place  $n$ , is admissible then we are done, by the same argument as in (5).

If the sequence  $i_0 \dots i_{n-k_1-1}100 \dots 0100 \dots$  is not admissible then we let  $i^{(2)} = i_0 \dots i_{n-k_1-1}00 \dots$  and proceed with  $i^{(2)}$  as above with  $i$ ; there is a maximal  $k_2 < n$  such that  $i_{k_2}^{(2)} \dots i_{n-1}^{(2)}0 = i_0^{(2)} \dots i_{n-k_2}^{(2)}$  and the two cylinders  $C_n^{\mathbb{P}}([i_0^{(2)} \dots i_{n-1}^{(2)}])$  and  $C_n^{\mathbb{P}}([i_0^{(2)} \dots i_{n-k_2-1}^{(2)}10 \dots 0])$  are both non-empty and neighbours.

This process is continued until we get an  $i^{(j)}$  with both  $i_0^{(j)} \dots i_{n-1}^{(j)}0 \dots$  and  $i_0^{(j)} \dots i_{n-1}^{(j)}100 \dots$  admissible. Note that  $0 < k_1 < k_2 < \dots < n$ , so

after at most  $n$  steps we end up with a cylinder  $C_n^{\mathbb{P}}([i_0^{(j)} \dots i_{n-1}^{(j)}])$  with  $|C_n^{\mathbb{P}}([i_0^{(j)} \dots i_{n-1}^{(j)}])| \geq \beta^{-n}$ . ■

**COROLLARY 1.** *For any  $\beta \in (1, \infty)$  and  $\varepsilon > 0$  there is a number  $n = n(\beta, \varepsilon)$  and a set  $V^{\mathbb{P}} \supset B(\beta, \varepsilon)$  such that  $V^{\mathbb{P}}$  can be written as a union of  $n$  cylinders of length  $n$  and  $|B(\beta, \varepsilon)| > (\beta + \varepsilon)^{-n-1}$ .*

*Proof.* Let  $\beta \in (1, \infty)$  and  $\varepsilon > 0$  be fixed. Put  $\beta' = \beta + \varepsilon$  and let  $n = \max\{k \in \mathbb{N} : \beta^{-k} > 2\varepsilon\}$ . Take  $V^{\mathbb{P}}$  to be a union of  $n$  consecutive cylinders of length  $n$ ,

$$V^{\mathbb{P}} = \bigcup_{i=1}^n C_{n,i}^{\mathbb{P}},$$

where  $\beta' \in C_{n,n}^{\mathbb{P}}$  and  $\alpha_i < \alpha_{i+1}$  if  $\alpha_i \in C_{n,i}^{\mathbb{P}}$  and  $\alpha_i \in C_{n,i+1}^{\mathbb{P}}$ . Lemma 3 implies that  $|V^{\mathbb{P}}| > \beta'^{-n} > 2\varepsilon = |B(\beta, \varepsilon)|$  and so  $V^{\mathbb{P}} \supset B(\beta, \varepsilon)$ .

Since  $n = \max\{k \in \mathbb{N} : \beta^{-k} > 2\varepsilon\}$  we have  $\beta'^{-n-1} < \beta^{-n-1} \leq 2\varepsilon = |B(\beta, \varepsilon)|$ . ■

It is easier to work in a subshift than in the parameter space of which  $D_\alpha(\beta_0, \beta_1)$  is a subset. Any subshift  $S_{\beta_2}$  of finite type has the property that the quantities

$$\frac{\text{the number of cylinders of length } n}{\beta_2^n}$$

and

$$\frac{\text{the diameter of a cylinder in } \mathcal{C}_n}{\beta_2^{-n}}$$

are bounded and bounded away from zero, uniformly over  $n$ . This makes the subshifts of finite type particularly easy to work with. Therefore we map the set  $D_\alpha(\beta_0, \beta_1)$  into a subshift of finite type  $S_{\beta_2}$  and estimate the dimension of the image. This is done in the following lemma. We will then transfer the estimate of the dimension of the image to an estimate on the dimension of  $D_\alpha(\beta_0, \beta_1)$ .

**LEMMA 4.** *Let  $\beta_0 < \beta_1 < \beta_2$  be such that  $\Xi(\beta_2)$  terminates with zeros. Then*

$$\dim_{\mathbb{H}}(\varrho_{\beta_2}^{-1}(D_\alpha(\beta_0, \beta_1))) \geq \frac{1}{1 + \alpha} \frac{\log \beta_1}{\log \beta_2}.$$

*Proof.* Let  $\underline{i} = \Xi(\beta_1)$  and take  $n$  so large that  $\beta_0 \notin U_n$ , where  $U_n = C_n^{\mathbb{P}}([i_0 \dots i_{n-1}])$  and  $(i_0 \dots i_{n-1})^\infty \in \Xi(U_n)$ . Let  $\beta_n \in (\beta_0, \beta_1) \setminus U_n$  be such that  $\Xi(\beta_n)$  terminates with zeros. Then the subshift  $S_{\beta_n}$  is of finite type (see [9]). Since  $\beta_1 \in U_n$  we have  $\beta_0 < \beta_n < \beta_1$ .

We will denote by  $\sigma^{-k}S_{\beta_n}$  the sets of sequences in  $S_2 = \Sigma_2 = \{0, 1\}^{\mathbb{N}}$  such that their  $k$ th shift is in  $S_{\beta_n}$ . That is,

$$\sigma^{-k}S_{\beta_n} = \{\underline{i} \in S_2 : \sigma^k(\underline{i}) \in S_{\beta_n}\}.$$

Then  $\Xi(U_n) \cap \sigma^{-n-1}S_{\beta_n} \subset K_{\beta_1}$ , where  $K_{\beta_1}$  is the set of sequences  $\underline{x} \in S_{\beta_1}$  such that  $\underline{x} = \Xi(\beta)$  for some  $\beta$ . Since  $(i_0 \dots i_{n-1})^\infty \in \Xi(U_n)$  we have  $\Xi(U_n) \cap \sigma^{-n-1}S_{\beta_n} \neq \emptyset$ .

Let  $\beta_2 > \beta_1$  be such that  $\underline{j} = \Xi(\beta_2)$  terminates with zeros. Then  $S_{\beta_2}$  is of finite type.

Take  $m$  so large that  $\sigma^m(\Xi(U_n)) \supset S_{\beta_1}$  and let  $\{k(l)\}_{l=0}^\infty$  be a sparse sequence such that  $k(0) > m$ . Let  $K_n^\alpha \subset S_{\beta_2}$  be the set of sequences  $\underline{x} \in \Xi(U_n) \cap \sigma^{-n-1}S_{\beta_n}$  with  $x_{k(l)} \dots x_{(1+\alpha)k(l)} = 0 \dots 0$  for every  $l \in \mathbb{N}$ .

We construct a measure  $\mu$  on  $S_{\beta_2}$  with support in  $K_n^\alpha$  in the following way. For any  $l \in \mathbb{N}$  define

$$S_l(\underline{x}) = C_{(1+\alpha)k(l)}([x_0x_1 \dots x_{k(l)-1} \underbrace{0 \dots 0}_{\alpha k(l) \text{ zeros}}], \beta_2).$$

Define the measure of  $S_l(\underline{x})$  by

$$\mu(S_0(\underline{x})) = \begin{cases} (\#\{S_0 : S_0 \cap K_n^\alpha \neq \emptyset\})^{-1} & \text{if } S_0(\underline{x}) \cap K_n^\alpha \neq \emptyset, \\ 0 & \text{otherwise.} \end{cases}$$

For  $l > 0$  we define  $\mu(S_l)$  recursively:

$$\mu(S_l(\underline{x})) = \begin{cases} \frac{\mu(S_{l-1}(\underline{x}))}{\#\{S_l(y) \subset S_{l-1}(\underline{x}) : S_l \cap K_n^\alpha \neq \emptyset\}} & \text{if } S_l(\underline{x}) \cap K_n^\alpha \neq \emptyset, \\ 0 & \text{otherwise.} \end{cases}$$

Since  $S_{\beta_n}$  is a subshift of finite type there is a constant  $c_2 > 0$  such that

$$c_2^{-1} \leq \frac{\#\{S_l(y) \subset S_{l-1}(\underline{x}) : S_l \cap K_n^\alpha \neq \emptyset\}}{\beta_n^{k(l)-(1+\alpha)k(l-1)}} \leq c_2$$

and since  $S_{\beta_2}$  is of finite type there is a constant  $c_3$  such that

$$c_3^{-1} \beta_2^{-(1+\alpha)k(l)} \leq |S_l| \leq c_3 \beta_2^{-(1+\alpha)k(l)}.$$

This implies that

$$\begin{aligned} \frac{\log \mu(S_l)}{\log |S_l|} &\geq \frac{\log \frac{\mu(S_{l-1})}{c_2 \beta_n^{k(l)-(1+\alpha)k(l-1)}}}{\log c_3^{-1} \beta_2^{-(1+\alpha)k(l)}} \\ &= \frac{\log \mu(S_{l-1}) - \log c_2 - k(l) \log \beta_n + (1 + \alpha)k(l-1) \log \beta_n}{-\log c_3 - (1 + \alpha)k(l) \log \beta_2} \\ &\rightarrow \frac{1}{1 + \alpha} \frac{\log \beta_n}{\log \beta_2}, \end{aligned}$$

as  $k(l)/k(l-1) \rightarrow \infty$ . This shows that

$$\dim_{\mathbb{H}}(\varrho_{\beta_2}^{-1}(D_\alpha(\beta_0, \beta_1))) \geq \dim_{\mathbb{H}} \mu \geq \frac{1}{1 + \alpha} \frac{\log \beta_n}{\log \beta_2}.$$

Let  $\beta_n \rightarrow \beta_1$ . ■

We are now ready to calculate the dimension of  $D_\alpha(\beta_0, \beta_1)$ .

**THEOREM 14.**  $\dim_{\mathbb{H}}(D_\alpha(\beta_0, \beta_1)) = 1/(1 + \alpha)$ .

*Proof.* Fix  $\beta_2 > \beta_1$  such that  $S_{\beta_2}$  is of finite type. Let  $\{U_k^{\mathbb{P}}\}$  be a  $\delta$ -cover of  $D_\alpha(\beta_0, \beta_1)$ . By Corollary 1 we can, for each  $k$ , find a set  $V_k^{\mathbb{P}} \supset U_k^{\mathbb{P}}$  such that  $|U_k^{\mathbb{P}}| > (\sup U_k^{\mathbb{P}})^{-n(k)-1}$  and  $V_k^{\mathbb{P}}$  is a union of  $n(k)$  cylinders of length  $n(k)$ ,

$$V_k^{\mathbb{P}} = \bigcup_{i=1}^{n(k)} C_{n(k),i}^{\mathbb{P}}.$$

To each  $V_k^{\mathbb{P}}$  there is a corresponding set  $V_k \subset S_{\beta_2}$ , defined in a natural way as the union of the corresponding cylinders in  $S_{\beta_2}$ . There is a constant  $c_1$  such that  $c_1 n(k) \beta_2^{-n(k)} \geq |V_k| \geq c_1^{-1} n(k) \beta_2^{-n(k)}$ . Since  $\{V_k^{\mathbb{P}}\}$  covers  $D_\alpha(\beta_0, \beta_1)$ , it follows that  $\{V_k\}$  covers  $\varrho_{\beta_2}^{-1}(D_\alpha(\beta_0, \beta_1))$ . Since  $\beta_2 > \beta_1$  there is a constant  $c_2$  such that  $|U_k^{\mathbb{P}}| > \beta_k'^{-n(k)-1} > c_2 n(k) \beta_2^{-n(k)} > c_2 c_1^{-1} |V_k|$ , where  $\beta_k' = \sup U_k^{\mathbb{P}}$ . This implies that

$$\begin{aligned} \sum |U_k^{\mathbb{P}}|^s &\geq (c_2 c_1^{-1})^s \sum |V_k|^s \\ &\geq (c_2 c_1^{-1})^s \inf \left\{ \sum |U_k|^s : \{U_k\} \text{ is a } (c_1 c_2^{-1} \delta)\text{-cover of} \right. \\ &\qquad \qquad \qquad \left. \varrho_{\beta_2}^{-1}(D_\alpha(\beta_0, \beta_1)) \right\}. \end{aligned}$$

But  $\{U_k^{\mathbb{P}}\}$  is an arbitrary  $\delta$ -cover of  $D_\alpha(\beta_0, \beta_1)$  so this proves that

$$\begin{aligned} \inf \left\{ \sum |U_k^{\mathbb{P}}|^s : \{U_k^{\mathbb{P}}\} \text{ is a } \delta\text{-cover of } D_\alpha(\beta_0, \beta_1) \right\} \\ \geq (c_2 c_1^{-1})^s \inf \left\{ \sum |U_k|^s : \{U_k\} \text{ is a } (c_1 c_2^{-1} \delta)\text{-cover of} \right. \\ \left. \varrho_{\beta_2}^{-1}(D_\alpha(\beta_0, \beta_1)) \right\}. \end{aligned}$$

We conclude that

$$\dim_{\mathbb{H}}(D_\alpha(\beta_0, \beta_1)) \geq \dim_{\mathbb{H}}(\varrho_{\beta_2}^{-1}(D_\alpha(\beta_0, \beta_1))) \geq \frac{1}{1 + \alpha} \frac{\log \beta_1}{\log \beta_2}$$

and let  $\beta_2 \rightarrow \beta_1$ . ■

**5.3. Convergence of dimensions in  $\beta$ -shifts.** The technique from the proof of Lemma 4 can be used to obtain the following theorem of convergence of dimensions for a general set.

**THEOREM 15 (Dimension Transference Principle).** *Let  $\beta_0 > 1$ . If  $E \subset S_{\beta_0}$  is such that for any  $\beta < \beta_0$  and any cylinder  $C_n$  of length  $n$  there exists an  $m = m(C_n)$  such that  $\dim_{\mathbb{H}}(E \cap S_\beta) = \dim_{\mathbb{H}}(C_n \cap \sigma^{-m}(E \cap S_\beta))$  and  $K_{\beta_0}$  is the set of  $\underline{x} \in S_{\beta_0}$  such that  $\underline{x} = \Xi(\beta)$  for some  $\beta < \beta_0$ ,*



then

$$\dim_{\mathbb{H}}(E \cap K_{\beta_0}) = \lim_{\beta \rightarrow \beta_0} \dim_{\mathbb{H}}(E \cap S_{\beta}),$$

where the dimensions are the dimensions as subsets of  $S_{\beta_0}$ .

*Proof.* We will first prove that  $\dim_{\mathbb{H}}(E \cap K_{\beta_0}) \leq \lim_{\beta \rightarrow \beta_0} \dim_{\mathbb{H}}(E \cap S_{\beta})$ . It is clear that  $K_{\beta_0} \subset \bigcup_{\beta < \beta_0} S_{\beta}$ . Hence, if  $\{\beta_n\}_{n=1}^{\infty}$  is an increasing sequence with  $\beta_n < \beta_0$  and  $\beta_n \rightarrow \beta_0$ , then

$$\begin{aligned} \dim_{\mathbb{H}}(E \cap K_{\beta_0}) &\leq \dim_{\mathbb{H}} \bigcup_{n=1}^{\infty} (E \cap S_{\beta_n}) \\ &= \sup_{n \geq 1} \dim_{\mathbb{H}}(E \cap S_{\beta_n}) = \lim_{n \rightarrow \infty} \dim_{\mathbb{H}}(E \cap S_{\beta_n}). \end{aligned}$$

It remains to show that  $\dim_{\mathbb{H}}(E \cap K_{\beta_0}) \geq \lim_{\beta \rightarrow \beta_0} \dim_{\mathbb{H}}(E \cap S_{\beta})$ . Fix  $\beta < \beta_0$ . We proceed as in the proof of Lemma 4 and take  $n$  so large that  $\beta \notin U_n := C_n^p(\beta_0) = C_n^p([i_0 \dots i_{n-1}])$  and  $(i_0 \dots i_{n-1})^{\infty} \in U_n$ . Then  $\Xi(U_n) \cap \sigma^{-m}(\Xi(U_n))(S_{\beta})$  is a non-empty subset of  $K_{\beta_0}$ , if we assume that  $m(\Xi(U_n)) > n$ . The required estimate follows from

$$\begin{aligned} \dim_{\mathbb{H}}(E \cap K_{\beta_0}) &\geq \dim_{\mathbb{H}}(E \cap (\Xi(U_n) \cap \sigma^{-m}(S_{\beta}))) \\ &= \dim_{\mathbb{H}}(\Xi(U_n) \cap \sigma^{-m}(E \cap S_{\beta})) = \dim_{\mathbb{H}}(E \cap S_{\beta}). \blacksquare \end{aligned}$$

### References

- [1] F. Blanchard,  *$\beta$ -expansions and symbolic dynamics*, Theor. Comput. Sci. 65 (1989), 131–141.
- [2] R. Bowen, *Topological entropy for noncompact sets*, Trans. Amer. Math. Soc. 184 (1973), 125–136.
- [3] J.-R. Chazottes, *Dimensions and waiting times for Gibbs measures*, J. Statist. Phys. 98 (2000), 305–320.
- [4] K. Falconer, *Fractal Geometry. Mathematical Foundations and Applications*, Wiley, Chichester, 1990.
- [5] A. Fan and J. Schmeling, *On fast Birkhoff averaging*, Math. Proc. Cambridge Philos. Soc. 135 (2003), 443–467.
- [6] A. Fan, J. Schmeling and S. Troubetzkoy, *Dynamical Diophantine Approximation*, arXiv:0705.4203v1.
- [7] V. Jarník, *Zur metrischen Theorie der diophantischen Approximationen*, Prace Mat.-Fiz. 1928–1929, 91–106.
- [8] A. Katok, *Lyapunov exponents, entropy and periodic orbits for diffeomorphisms*, Inst. Hautes Études Sci. Publ. Math. 51 (1980), 137–173.
- [9] W. Parry, *On the  $\beta$ -expansion of real numbers*, Acta Math. Acad. Sci. Hungar. 11 (1960), 401–416.
- [10] W. Parry and M. Pollicott, *Zeta functions and the periodic orbit structure of hyperbolic dynamics*, Astérisque 187-188 (1990).
- [11] Ya. Pesin, *Dimension Theory in Dynamical Systems. Contemporary Views and Applications*, Chicago Lectures in Math., Univ. of Chicago Press, Chicago, 1997.

- [12] Ya. Pesin and B. S. Pitskel', *Topological pressure and the variational principle for noncompact sets*, Functional Anal. Appl. 18 (1984), 307–318.
- [13] J. Schmeling, *Symbolic dynamics for  $\beta$ -shifts and self-normal numbers*, Ergodic Theory Dynam. Systems 17 (1997), 675–694.
- [14] P. Walters, *An Introduction to Ergodic Theory*, Grad. Texts in Math. 79, Springer, New York, 1982.

Centre for Mathematical Sciences

Lund University

Box 118

SE-221 00 Lund, Sweden

E-mail: tomasp@maths.lth.se

joerg@maths.lth.se

<http://www.maths.lth.se/matematiklth/personal/tomasp/>

<http://www.maths.lth.se/matematiklth/personal/joerg/>

*Received on 23.10.2006  
and in revised form on 20.9.2007*

(5304)