TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 363, Number 8, August 2011, Pages 4421–4442
S 0002-9947(2011)05357-2
Article electronically published on March 4, 2011

AN ANALOGUE OF COBHAM'S THEOREM FOR FRACTALS

BORIS ADAMCZEWSKI AND JASON BELL

ABSTRACT. We introduce the notion of k-*self-similarity* for compact subsets of $\mathbb{R}^n$ and show that it is a natural analogue of the notion of k-automatic subsets of integers. We show that various well-known fractals such as the triadic Cantor set, the Sierpiński carpet or the Menger sponge turn out to be k-self-similar for some integers k. We then prove an analogue of Cobham's theorem for compact sets of $\mathbb{R}$ that are self-similar with respect to two multiplicatively independent bases k and ℓ. Namely, we show that X is both a k- and an ℓ-self-similar compact subset of $\mathbb{R}$ if and only if it is a finite union of closed intervals with rational endpoints.

1. Introduction

The notion of self-similarity is fundamental in the study of fractals. We recall (see Falconer [12]) that a compact topological space X is *self-similar* if there is a finite set of non-surjective homeomorphisms $f_1, \ldots, f_n : X \to X$ such that

$$X = \bigcup_{i=1}^{n} f_i(X).$$

It can be motivated by looking at the usual triadic Cantor set C, which is the closed subset of $[0,1]$ consisting of all numbers whose ternary expansion does not contain any 1s. We note that

$$C = \frac{1}{3}C \cup \left(\frac{1}{3}C + \frac{2}{3}\right).$$

The fact that C is a disjoint union of a finite number of images of itself under affine transformations tells us that it is self-similar.

With this in mind, we define the notion of a k-kernel for subsets of $[0,1]^d$. The k-kernel essentially looks at the possible sets one can obtain by taking the intersection of X with certain cubes in $[0,1]^d$ with side length $1/k^a$ for some positive integer a and then scaling by a factor of k^a.

Received by the editors July 5, 2009 and, in revised form, January 28, 2010 and March 23, 2010.

2010 *Mathematics Subject Classification.* Primary 28A80, 11B85.

The first author was supported by the ANR through the project "DyCoNum"–JCJC06 134288. He also thanks Jean-Paul Allouche for pointing out relevant references.

The second author thanks NSERC for its generous support.

Definition 1.1. Given a subset $X \subseteq [0,1]^d$, we define the *k-kernel* to be the collection of distinct subsets of the form

$$\left\{ (k^a x_1 - b_1, \ldots, k^a x_d - b_d) \in [0,1]^d \ : \ (x_1, \ldots, x_d) \in X \cap \prod_{j=1}^{d} [b_j/k^a, (b_j+1)/k^a] \right\},$$

where $a \geq 0$ and $0 \leq b_1, \ldots, b_d < k^a$ are integers.

We then define the notion of k-self-similarity in terms of k-kernels.

Definition 1.2. A compact set $X \subseteq [0,1]^d$ is said to be *k-self-similar* if it has a finite k-kernel.

As we will see in Section 2, many famous examples of fractals are in fact k-self-similar sets for some k. For instance, the Cantor set is a 3-self-similar subset of $\mathbb{R}$, the Sierpiński carpet and the Reverend Back's abbey floor are 3-self-similar subsets of $\mathbb{R}^2$, Pascal's triangle modulo 2 is a 2-self-similar subset of $\mathbb{R}^2$, and the Menger sponge is a 3-self-similar subset of $\mathbb{R}^3$.

Our definition of self-similar compact subsets of $\mathbb{R}^n$ is given in terms of kernels. Actually, there are famous subsets of integers that can be defined in a similar way: *automatic* or *recognizable sets of integers*. We now briefly describe this analogy. We refer to the book of Allouche and Shallit [6] for a more formal and complete introduction to this topic.

Let $k \geq 2$ be a natural number. A set $\mathcal{N} \subset \mathbb{N}$ is said to be *k-automatic* if there is a finite-state machine that accepts as input the expansion of n in base k and outputs 1 if $n \in \mathcal{N}$ and 0 otherwise. For example, the set of Thue–Morse integers $1, 2, 4, 7, 8, 11, 13, \ldots$, formed by the integers whose sum of binary digits is odd, is 2-automatic. The associated automaton is given in Figure 1. It has two states. This automaton successively reads the binary digits of n (starting, say, from the most significant digit and the initial state q_0) and thus ends the reading either in state q_0 or in state q_1. The initial state q_0 gives the output 0, while q_1 gives the output 1.

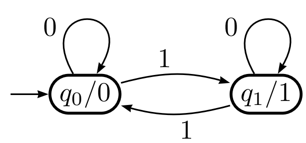

FIGURE 1. The finite-state automaton recognizing the set of Thue–Morse integers.

An equivalent formulation of k-automatic sets of integers is given in terms of *k-kernels*. Given a set $\mathcal{N} \subseteq \mathbb{N}$, the *k-kernel* of $\mathcal{N}$ is defined as the collection of distinct sets

$$\{K_{a,b} \ : \ a \geq 0; \ 0 \leq b < k^a\},$$

where

$$K_{a,b} = \{n \geq 0 \mid k^a n + b \in \mathcal{N}\}.$$

Then we have the following characterization [6]: a set $\mathcal{N} \subseteq \mathbb{N}$ is k-automatic if and only if its k-kernel is finite.

Another typical 2-automatic set of integers is given by the powers of 2: $1, 2, 4, 8, 16, \ldots$. Though these integers have very simple expansion in base 2, one can observe that this is not the case when writing them in base 3. One of the most important results in the theory of automatic sets formalizes this idea. Recall that two integers k and l larger than 1 are multiplicatively independent if $\log(k)/\log(l) \notin \mathbb{Q}$. Then Cobham's theorem says that only very well-behaved sets of integers can be automatic with respect to two multiplicatively independent numbers [9]. In 1969, Cobham proved the following result.

Theorem 1.3 (Cobham)**.** *Let k and ℓ be two multiplicatively independent integers. Then a set $\mathcal{N} \subseteq \mathbb{N}$ is both k- and l-automatic if and only if it is the union of a finite set and a finite number of arithmetic progressions.*

Our main result is an analogue of Cobham's theorem for k-self-similar compact subsets of $\mathbb{R}$.

Theorem 1.4. *Let k and ℓ be two multiplicatively independent natural numbers. Then a compact set $X \subseteq [0, 1]$ is both k- and ℓ-self-similar if and only if it is a finite union of closed intervals with rational endpoints.*

Thus the triadic Cantor set cannot be 2- or 7-self-similar. Theorem 1.4 is only concerned with one-dimensional sets, but we expect a similar picture in every dimension. More precisely, we suggest the following multi-dimensional version of Theorem 1.4.

Conjecture 1.5. *Let k and ℓ be two multiplicatively independent natural numbers. Then a compact set $X \subseteq [0, 1]^d$ is both k- and ℓ-self-similar if and only if it is a finite union of polyhedra whose vertices have rational coordinates.*

Notice that Cobham's theorem has been generalized to subsets of $\mathbb{N}^d$ by Semenov [23]. Conjecture 1.5 can thus be thought of as an analogue of Semenov's result.

To end this introduction, we mention that similar "independence principles" with respect to two multiplicatively independent integers are also expected in other contexts. This is a source of difficult questions arising from various fields. As an illustration, we quote below three famous open problems that rest on such a principle. A long-standing question in dynamical systems is the so-called $\times 2 \times 3$ problem addressed by Furstenberg [13]: given two multiplicatively independent integers $k, l \geq 2$, prove that the only Borel measures on $[0, 1]$ that are simultaneously ergodic for $T_k(x) = kx \pmod 1$ and $T_l(x) = lx \pmod 1$ are the Lebesgue measure and measures supported by those orbits that are periodic for both actions T_k and T_l. With a number-theoretic flavor, we recall another problem attributed to Mahler and Mendès France (see, for instance, Adamczewski and Bugeaud [1]): given a binary sequence $(a_n)_{n\geq 0} \in \{0, 1\}^{\mathbb{N}}$, prove that

$$\sum_{n\geq 0} \frac{a_n}{2^n} \quad \text{and} \quad \sum_{n\geq 0} \frac{a_n}{3^n}$$

are both algebraic numbers only if both are rational numbers. The last problem we mention appeared in work of Ramanujan (see Waldschmidt [26]): prove that there is no irrational real number x such that both 2^x and 3^x are integers. This corresponds to a particular instance of the four exponentials conjecture, a famous open problem in transcendence theory [25, Chapter 1, p. 15].

The outline of this paper is as follows. In Section 2, we show how one can simply associate a fractal set with a finite automaton. Our approach is inspired by recent work of Kedlaya concerning an extension of Christol's theorem to Hahn's power series [17]. These fractal sets are termed automatic fractals. We give several examples of famous fractals that turn out to be automatic and prove that automatic fractals all are self-similar sets. We then prove Theorem 1.4 in Section 3. We note that our proof of Theorem 1.4 does not rely on Cobham's theorem. In Section 4, we investigate the link between the Hausdorff dimension of automatic fractals and the entropy of languages naturally associated with automatic fractals. We remark that some results of Mauldin and Williams [18] should allow one to compute the Hausdorff dimension of automatic fractals. We also observe that Hartmanis and Stearns [16] and Barbé and von Haeseler [7] previously considered similar fractals in the framework of automata.

2. Automatic fractals

Finite automata are devices that accept finite words as input. This can be naturally used to recognize sets of integers, since numbers correspond to finite words when representing them in an integer base. In contrast, most real numbers have infinite expansions in integer bases and are thus related to infinite words. Hence it is unclear how to properly define a notion of subsets of real numbers recognized by finite automata. In this section, we show how one can simply associate fractal sets with finite automata, called *automatic fractals*. We give several examples of famous automatic fractals and prove that a k-automatic fractal is always a k-self-similar set.

2.1. **Automatic sets revisited.**

Kedlaya [17] introduced the notion of automaticity for subsets of k-adic rational numbers by adding a radix point to the input set for a finite-state automaton. The key point is that k-adic rationals are exactly those real numbers with a finite expansion in base k. Such real numbers are represented by finite words and can thus be read by a finite k-automaton. This is the way in which we will extend the notion of automaticity beyond subsets of integers.

Let $k > 1$ be a positive integer and $\Sigma_k := \{0, 1, \ldots, k-1\}$. Given a natural number n and a positive integer $k \geq 2$, we let $[n]_k$ denote the base-k expansion of n and we let S_k denote the set of non-negative k-adic rationals; i.e.,

$$S_k \ = \ \{a/k^b \mid a, b \in \mathbb{Z},\ a \geq 0\}. \tag{1}$$

Then a k-adic rational number has a finite base-k expansion of the form $[n]_k \bullet [m]_k$, where $\bullet$ is the radix point.

We set

$$\Sigma'_k \ = \ \{0, 1, \ldots, k-1, \bullet, l\},$$

and we let $\mathcal{L}(k)$ denote the language over the alphabet Σ'_k consisting of all words over Σ'_k with exactly one occurrence of the letter '$\bullet$' (the radix point) and whose first and last letters are not equal to 0. (We note that the fact that we exclude strings whose initial and terminal letters are 0 means we have the awkward looking expression $[\bullet]_k = 0$.) This is a regular language [17, Lemma 2.3.3]. We note that there is a bijection $[\,\cdot\,]_k : \mathcal{L}(k) \to S_k$ given by

$$s_1 \cdots s_{i-1} {\bullet} s_{i+1} \cdots s_n \in \mathcal{L}(k) \ \mapsto \ \sum_{j=1}^{i-1} s_j k^{i-1-j} + \sum_{j=i+1}^{n} s_j k^{i-j},$$

where $s_1, \ldots, s_{i-1}, s_{i+1}, \ldots, s_n \in \{0, 1, \ldots, k-1\}$.

So, for example, we have $[110{\bullet}32]_4 = [20{\bullet}875]_{10} = 167/8$. We now recall the definition of a k-automatic subset of S_k.

Definition 2.1. We say that a subset S of S_k is k-*automatic* if there is a finite-state automaton that takes words over Σ_k' as input and has the property that a word $W \in \mathcal{L}_k$ is accepted by the automaton if and only if $[W]_k \in S$.

More generally, we can define automatic subsets of S_k^d, mimicking the construction of Salon [22]. For a natural number $d \geq 1$, we create the alphabet $\Sigma_k'(d)$ to be the alphabet $(\Sigma_k')^d$ consisting of all d-tuples of elements of Σ_k'. Then an element of S_k^d is simply a d-tuple of elements of S_k. With this in mind, we construct a regular language $\mathcal{L}_k(d) \subseteq (\Sigma_k'(d))^*$ as follows. Given a k-adic rational $a \in S_k$, we can write it uniquely as

$$a \ = \ \sum_{j=-\infty}^{\infty} e_j(a) k^j,$$

in which $e_j(a) \in \{0, \ldots, k-1\}$, and there is some natural number N, depending on a, such that $e_j(a) = 0$ whenever $|j| > N$. Let $(a_1, \ldots, a_d)$ be a d-tuple of non-negative k-adic rationals and let

$$h := \max\{j \ : \ \text{there exists some } i \text{ such that } e_j(a_i) \neq 0\} \cup \{-1\}.$$

Similarly, we let

$$\ell := \min\{j \ : \ \text{there exists some } i \text{ such that } e_j(a_i) \neq 0\} \cup \{1\}.$$

We can then produce an element

$$w_k(a_1, \ldots, a_d) := (w_1, \ldots, w_d) \in (\Sigma_k(d))^*$$

corresponding to $(a_1, \ldots, a_d)$ by defining

$$w_i := e_h(a_i) e_{h-1}(a_i) \cdots e_0(a_i) {\bullet} e_{-1}(a_i) \cdots e_\ell(a_i).$$

In other words, we are taking the base k-expansions of $a_1, \ldots, a_d$ and then "padding" the expansions of each a_i at the beginning and the end with 0s if necessary to ensure that each expansion has the same length pre-radix and post-radix length. For example, if $d = 2$ and $k = 3$, then $w_3(14, 3) = (112{\bullet}, 010{\bullet})$ and $w_3(1/3, 1/9) = ({\bullet}10, {\bullet}01)$. We then take $\mathcal{L}_k(d)$ to be the collection of words of the form

$$w_k(a_1, a_2, \ldots, a_d),$$

where $(a_1, a_2, \ldots, a_d) \in S_k^d$. Then there is an obvious way to extend the map $[\cdot]_k$ to a bijection $[\,\cdot\,]_k : \mathcal{L}_k(d) \to S_k^d$; namely,

$$[w_k(a_1, \ldots, a_d)]_k \ := \ (a_1, \ldots, a_d).$$

Definition 2.2. We say that a subset S of S_k^d is k-*automatic* if there is a finite-state automaton that takes words over $\mathcal{L}_k(d)$ as input and has the property that a word $W \in \mathcal{L}_k(d)$ is accepted by the automaton if and only if $[W]_k \in S$.

We will also use the notion of a k-automatic function from S_k^d to a finite set.

Definition 2.3. Let Δ be a finite set. We say that a function $f : S_k^d \to \Delta$ is k-*automatic* if there is a finite-state automaton that takes words over $\mathcal{L}_k(d)$ as input and has the property that, reading a word $W \in \mathcal{L}_k(d)$, the automaton outputs $f([W]_k)$.

We make the following remark, which is a translation in our context of Theorem 6.6.2 of Allouche and Shallit [6].

Remark 2.4. If a subset S of S_k^d is k-*automatic*, then there are only finitely many distinct subsets of S_k^d of the form

$$\left\{(x_1, \ldots, x_d) \in [0,1]^d \mid (x_1/k^a + b_1/k^a, \ldots, x_d/k^a + b_d/k^a) \in S\right\},$$

where $a \geq 0$ and $0 \leq b_1, \ldots, b_d < k^a$.

The set S_k^d is countable, and hence automatic sets of S_k^d are not very interesting in the framework of fractals. We can actually overcome this deficiency by considering the closure of automatic sets. As an example, we look at the Cantor set C. Since it can be described as the set of all $x \in [0,1]$ which have a ternary expansion that does not have any 1s, the set $C \cap S_3$ is just the set of 3-adic rationals in $[0,1]$ whose ternary expansion does not contain any 1s. We note that this set is a 3-automatic subset of S_3, as it can be described as being all numbers in S_3 that begin with a radix point and do not contain a 1. We also observe that $C \cap S_3$ is dense in C and hence the Cantor set can be realized as the closure of a 3-automatic subset of S_3. This brings us to our definition of k-automatic fractals.

Definition 2.5. We say that a compact subset $X \subseteq \mathbb{R}^d$ is a k-automatic fractal if it is the closure of a k-automatic subset of S_k^d.

2.2. **Examples of famous automatic fractals.** To illustrate the relevance of the notion of automatic fractals, we give several examples of classical fractals that turn out to be automatic.

Example 2.6. The Cantor set is a 3-automatic fractal (see Figure 3). It is the closure of the following 3-automatic set of S_3:

$$\{(\bullet[n_1 n_2 \cdots n_k]_3) \mid n_i \neq 1 \; \forall i \in [1,k]\}.$$

It is associated with the following automaton (see Figure 2).

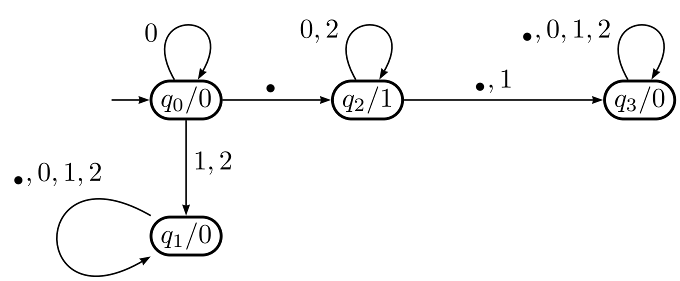

FIGURE 2. The finite automaton associated with the triadic Cantor set.

Figure 3. The first steps in the construction of the triadic Cantor set.

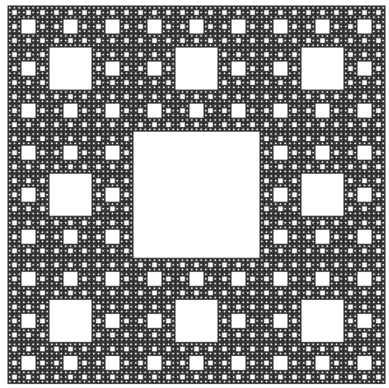

Figure 4. The Sierpiński carpet.

Example 2.7. The Sierpiński carpet is a 3-automatic fractal of $\mathbb{R}^2$ (see Figure 4).

It can be defined as the set of pairs of real numbers (x, y) in $[0,1]^2$ such that for every positive integer n, the n-th digit of the ternary expansion of x and of y are not both equal to 1. This set is thus the closure of the 3-automatic set of S_3^2 defined by

$$\{({}_\bullet[n_1 n_2 \cdots n_k]_3, {}_\bullet[m_1 m_2 \cdots m_l]_3) \mid (n_i, m_i) \neq (1,1)\ \forall i \in [1, \min(k,l)]\}.$$

Example 2.8. The Menger sponge (see Figure 5) is defined as a three-dimensional analogue of the Sierpiński carpet (see Addison [3, Section 2.10]). It is a 3-automatic fractal of $\mathbb{R}^3$. It can be defined as the set of 3-tuples of real numbers (x, y, z) in $[0,1]^3$ such that for every positive integer n, at most one of the n-th digit of the ternary expansion of x, the n-th digit of the ternary expansion of y, and the n-th digit of the ternary expansion of z is equal to 1.

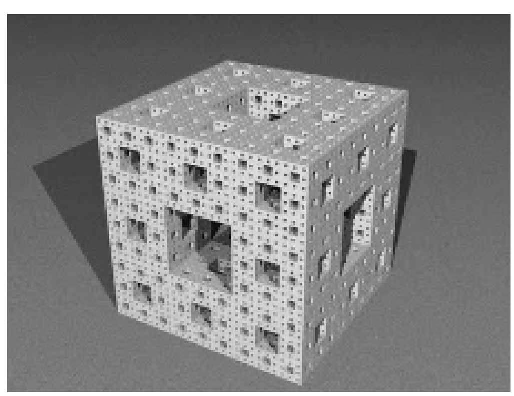

Figure 5. The Menger sponge.

Example 2.9. Pascal's triangle modulo 2 (see Figure 6) is the subset of $[0,1]^2$ formed by taking the closure of

$$\left\{ ({}_\bullet[n]_2, {}_\bullet[m]_2) \mid \binom{m}{n} \equiv 1 \bmod 2 \right\}.$$

We infer that this set is a 2-automatic fractal of $\mathbb{R}^2$ [6, p. 420].

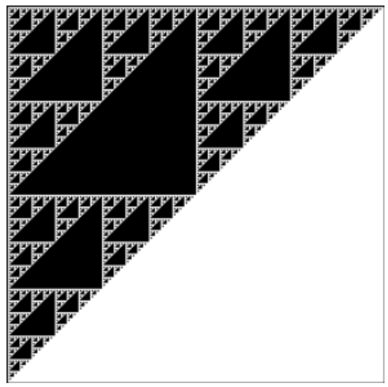

FIGURE 6. Pascal's triangle modulo 2.

Example 2.10. We give the following example of a 5-automatic fractal of $\mathbb{R}^2$ (see Figure 7).

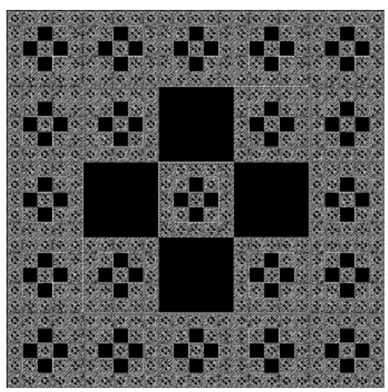

FIGURE 7. An example of a 5-automatic fractal of $\mathbb{R}^2$.

It is obtained as the closure of the set

$$\{ ({}_\bullet[n_1 n_2 \cdots n_k]_5, {}_\bullet[m_1 m_2 \cdots m_l]_5) \mid (n_i, m_i) \notin \{(1,2),(2,1),(2,3),(3,2)\}, \quad \forall i \in [1, \min(k,l)] \}.$$

The above examples of automatic fractals are all related to simple automata and should only be considered as basic examples. There is actually a large variety of finite automata, and more involved ones can be used to construct fractals with a much more complex structure (see, for instance, Example 2.11). On the other hand, some non-trivial automata give rise to trivial fractals. For instance, we can use a small variation of the Thue–Morse automaton represented in Figure 1 to define a finite 2-automaton that recognizes exactly the elements of $S_2 \cap [0,1]$ whose binary expansion contains an odd number of 1s. Though this set is not trivial, its closure is the whole interval $[0,1]$.

Example 2.11. Reverend Back's abbey floor is a 3-automatic fractal of $\mathbb{R}^2$ (see Wegner [27] and also Allouche and Shallit [6, page 410] for a definition).

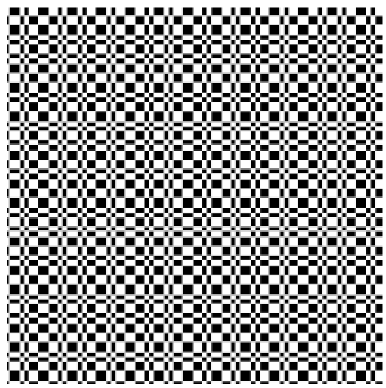

FIGURE 8. Reverend Back's abbey floor.

2.3. **Automatic fractals and self-similarity.** The following proposition gives the relationship between k-automatic fractals and k-self-similar sets.

Proposition 2.12. *If a compact subset X of $[0,1]^d$ is a k-automatic fractal, then it is k-self-similar.*

Proof. Suppose that $X \subseteq [0,1]^d$ is a k-automatic fractal. Then there is a k-automatic subset C of S_k^d whose closure is X. Then, by Remark 2.4, there are only finitely many distinct sets of the form

$$\{(x_1,\ldots,x_d) \in S_k^d \cap [0,1]^d \mid (x_1/k^a + b_1/k^a, \ldots, x_d/k^a + b_d/k^a) \ \in C\}$$

with $a \geq 0$ and $b_1, b_2, \ldots, b_d \in [1, k^a]$. Let $C_1, \ldots, C_m$ denote these sets and let $X_i \subseteq [0,1]^d$ denote the closure of C_i.

We claim that X has k-kernel $\{X_1, \ldots, X_m\}$ and thus is k-self-similar. To see this, let us fix a non-negative integer a and some integers $b_1, b_2, \ldots, b_d \in [0, k^a]$. Thus there exists an integer i, $1 \leq i \leq m$, such that

$$\{(x_1,\ldots,x_d) \in S_k^d \cap [0,1]^d \mid (x_1/k^a + b_1/k^a, \ldots, x_d/k^a + b_d/k^a) \in C\} \ = \ C_i.$$

It then follows from the definition of X_i that

$$\overline{\{(x_1,\ldots,x_d) \in S_k^d \cap [0,1]^d \mid (x_1/k^a + b_1/k^a, \ldots, x_d/k^a + b_d/k^a) \in C\}} = \overline{C_i} = X_i,$$

which is equivalent to

$$\{(x_1,\ldots,x_d) \in [0,1]^d \mid (x_1/k^a + b_1/k^a, \ldots, x_d/k^a + b_d/k^a) \in X\} = X_i.$$

Consequently

$$\Big\{(k^a x_1 - b_1, \ldots, k^a x_d - b_d) \in [0,1]^d \ : \ (x_1, \ldots, x_d) \in X \cap \prod_{j=1}^{d} [b_j/k^a, (b_j+1)/k^a] \Big\} = X_i.$$

This concludes the proof since there are only a finite number of possible choices for i. □

We note that the converse of Proposition 2.12 does not hold in general. This is not difficult to see: since there are only countably many finite-state automata with input alphabet Σ_k', there are only countably many one-dimensional k-automatic fractals. On the other hand, for each $\alpha \in (1/2, 1]$, we have that the set $X_\alpha := \{\alpha/k^m \mid m \geq 1\} \cup \{0\}$ is k-self-similar. Hence there are uncountably many k-self-similar subsets of $[0,1]$. We note that for $\alpha \notin S_k$, X_α does not even intersect the set S_k.

3. Proof of Theorem 1.4

In this section we prove Theorem 1.4. The main idea we use is that if $X \subset [0,1]$ is compact and is not a finite union of closed intervals, then its complement is a countable union of disjoint open intervals and the endpoints of these intervals have a limit point in $[0,1]$.

We first prove the easier part of Theorem 1.4. That is, a finite union of closed intervals with rational endpoints is always k-self-similar for every integer $k \geq 2$. This result follows from Propositions 3.1 and 3.2 below.

Proposition 3.1. *Suppose* $X = \bigcup_{i=1}^{m} [a_i, b_i] \subseteq [0,1]$ *is k-self-similar and the intervals*

$$[a_1, b_1], \ldots, [a_m, b_m]$$

are disjoint. Then $a_i, b_i \in \mathbb{Q}$ *for* $1 \leq i \leq m$.

Proof. Let $X = X_1, \ldots, X_d$ be the k-kernel of X. Then each X_i is a finite disjoint union of closed intervals. Let $\mathcal{S}$ denote the set of endpoints of the closed intervals that make up these sets. Then $\mathcal{S}$ is a finite set. Suppose that $x \in \mathcal{S}$ and let ${\scriptstyle\bullet} a_1 a_2 a_3 \cdots$ denote the base-k expansion of x. By definition of the k-kernel of X, we have that ${\scriptstyle\bullet} a_i a_{i+1} \cdots \in \mathcal{S}$ for every positive integer i. Since $\mathcal{S}$ is finite, we see that

$${\scriptstyle\bullet} a_i a_{i+1} a_{i+2} \cdots \ = \ {\scriptstyle\bullet} a_j a_{j+1} a_{j+2} \cdots$$

for some positive integers i and j with $j > i$. It follows that x is a rational number, concluding the proof. □

Proposition 3.2. *Suppose* $X \subseteq [0,1]$ *is a finite union of closed intervals with rational endpoints. Then X is k-self-similar for every integer* $k \geq 2$.

Proof. Let $k \geq 2$ be an integer. Let us assume that $X = [a_1, b_1] \cup \cdots \cup [a_d, b_d]$ with $b_i < a_{i+1}$ and $(a_i, b_i) \in \mathbb{Q}^2$ for $1 \leq i < d$. Then there exists a positive integer n such that

$$a_i k^n (k^n - 1), b_i k^n (k^n - 1) \in \mathbb{N}$$

for $1 \leq i \leq d$. Let

$$S = \{a/k^n(k^n - 1) \mid 0 \leq a \leq k^n(k^n - 1)\},$$

and let $\mathcal{T}$ denote the collection of subsets of $[0, 1]$ that can be written as a finite union of closed intervals with endpoints all in S. Then $\mathcal{T}$ is finite. Notice that $X \in \mathcal{T}$ and

$$\{k^i x - j \mid x \in [j/k^i, (j + 1)/k^i] \cap X\} \in \mathcal{T}$$

for every positive integer i and every non-negative integer $j \in [0, k^i)$. Thus the k-kernel of X is finite, and so X is k-self-similar. □

We now introduce the notion of a *wild point* that will often be used in the sequel.

Definition 3.3. Let $X \subseteq [0, 1]$. We say that a point $\beta \in X$ is a *right wild point* of X if β is a limit point of both $X \cap (\beta, 1]$ and $X^c \cap (\beta, 1]$. We say β is a *left wild point* of X if it is a limit point of both $X \cap [0, \beta)$ and $X^c \cap [0, \beta)$. We simply say β is a *wild point* of X if it is either a left or a right wild point of X.

For example, if $C \subseteq [0, 1]$ denotes the triadic Cantor set, then $1/3$ is a left wild point of C but is not a right wild point.

Lemma 3.4. *Suppose that $X \subseteq [0, 1]$ is closed and $T : [0, 1] \to [0, 1]$ is a continuous injective map satisfying $T(X) \subseteq X$ and $T(X^c) \subseteq X^c$. Then:*

(1) *If $\beta \in X$ is a wild point of X, then $T(\beta)$ is a wild point of X.*
(2) *If β_n are wild points of X and $\beta_n \to \beta$, then β is a wild point of X.*

Proof. The proof is a straightforward consequence of the definition of a wild point. □

For our next lemma we need to introduce some notation. Given a non-zero element $\alpha \in S_k$, let $v_k(\alpha)$ denote the k-adic valuation of α, that is, the largest integer n such that $k^{-n}\alpha \in \mathbb{Z}$. When $\alpha = 0$, our convention is that $v_k(\alpha) = 0$. Thus v_k does not fit exactly with the usual definition of the k-adic valuation for which we should have $v_k(0) = +\infty$. Lemma 3.5 shows a connection between k-self-similarity and the k-automatic function defined over S_k.

Lemma 3.5. *Let $k \geq 2$ be an integer, α an element of S_k and set $v(\alpha) := v_k(\alpha)$. Suppose that $Y \subseteq [0, 1]$ is k-self-similar. Then there are distinct sets $Y = Y_1, \ldots, Y_m$ and a k-automatic function $f : S_k \cap [0, 1) \to \{1, 2, \ldots, m\}$ such that for $\alpha \in S_k \cap [0, 1)$ we have*

$$Y_{f(\alpha)} \ = \ \{k^{-v(\alpha)}(x - \alpha) \mid x \in [\alpha, \alpha + k^{v(\alpha)}] \cap Y\}.$$

In particular, if $\gamma \in \mathbb{Q} \cap (0, 1)$ and

$$a_n = \lfloor k^n \gamma \rfloor,$$

then the sequence of sets

$$T_n \ := \ \{k^n x - a_n \mid x \in [a_n/k^n, (a_n + 1)/k^n] \cap Y\}$$

is eventually periodic.

Proof. Since Y is a k-self-similar set, there are only a finite number of distinct sets of the form

$$\{k^{-v(\alpha)}(x - \alpha) \mid x \in [\alpha, \alpha + k^{v(\alpha)}] \cap Y\},$$

with $\alpha \in S_k \cap [0,1]$. Let $Y = Y_1, Y_2, \ldots, Y_m$ denote these sets (the set Y corresponds to the case $\alpha = 0$ since by convention $v_k(0) = 0$). Since these sets are distinct, we first note that one can define a map $f : S_k \cap [0,1) \to \{1, 2, \ldots, m\}$ such that for every $\alpha \in S_k \cap [0,1)$, we have $f(\alpha) = j$ if $\{k^{-v(\alpha)}(x-\alpha) \mid x \in [\alpha, \alpha + k^{v(\alpha)}] \cap Y\} \ = \ Y_j$.

Recall that the k-kernel of a function whose domain is S_k is defined in an analogous manner to how k-kernels of functions whose domain is $\mathbb{N}$ is defined. Namely, if $f : S_k \to \Delta$ is a map, the k-kernel of f is the collection of all functions of the form $g(x) = f((x+c)/k^a)$ with $a \geq 0$ and $0 \leq c < k^a$. From what is known about automatic functions whose domain is $\mathbb{N}$ (see Allouche and Shallit [6]), it is easily checked that having a finite k-kernel is the same as being k-automatic for such a function f. To prove the first part of Lemma 3.5, it is thus sufficient to show that the k-kernel of the function f is finite.

We note that for $\alpha \in S_k \cap (0,1)$, the set

$$\{k^{-v(\alpha)}(x-\alpha) \mid x \in [\alpha, \alpha + k^{v(\alpha)}] \cap Y_j\}$$

must be one of $Y_1, \ldots, Y_m$. The reason for this is that there is some $\beta \in S_k \cap [0,1)$ such that

$$Y_j = \{k^{-v(\beta)}(x-\beta) \mid x \in [\beta, \beta + k^{v(\beta)}] \cap Y\}.$$

Then it follows that

$$\begin{aligned}
&\{k^{-v(\alpha)}(x-\alpha) \mid x \in [\alpha, \alpha + k^{v(\alpha)}] \cap Y_j\} = \\
&\qquad = \{k^{-v(\alpha)-v(\beta)}(x-\beta') \mid x \in [\beta', \beta' + k^{v(\alpha)+v(\beta)}] \cap Y\},
\end{aligned}$$

where $\beta' = \beta + k^{v(\beta)}\alpha$. For every positive integer $j \in [1,m]$, let f_j denote the function defined by

$$Y_{f_j(\alpha)} \ = \ \{k^{-v(\alpha)}(x-\alpha) \mid x \in [\alpha, \alpha + k^{v(\alpha)}] \cap Y_j\}$$

for $\alpha \in S_k \cap [0,1)$. Let $\alpha \in S_k \cap [0,1)$. Let $i \in \{0, 1, 2, \ldots, k-1\}$ and set $\ell = f_j(i/k)$. Then

$$\begin{aligned}
Y_{f_j((i+\alpha)/k)} &= \{k^{1-v(\alpha)}(x - i/k - \alpha/k) \mid x \in [i/k + \alpha/k, i/k + \alpha/k + k^{v(\alpha)-1}] \cap Y_j\} \\
&= \{k^{-v(\alpha)}(x-\alpha) \mid x \in [\alpha, \alpha + k^{v(\alpha)}] \cap Y_\ell\} \\
&= Y_{f_\ell(\alpha)}.
\end{aligned}$$

Since $f = f_1$, it follows that the k-kernel of f is contained in $\{f_1, \ldots, f_m\}$ and hence f is k-automatic.

For the second part of Lemma 3.5, note that T_n is simply $Y_{f(a_n/k^n)}$. We thus have to prove that the sequence $(f(a_n/k^n))_{n\geq 1}$ is eventually periodic. Since γ is rational, its expansion in base k is eventually periodic. Thus there exist two finite words $U = a_1a_2\cdots a_r$ (possibly empty) and $V = a_{r+1}a_{r+2}\cdots a_{r+s}$ such that

$$\gamma = [0_\bullet UV^\infty]_k = [0_\bullet UVVV\cdots]_k.$$

Then $a_n = \lfloor k^n\gamma \rfloor = [a_1a_2\cdots a_n{}_\bullet]_k$ and thus

$$a_n/k^n = [0_\bullet a_1a_2\cdots a_n]_k.$$

On the other hand, an easy adaptation of Theorem 5.5.2 from Allouche and Shallit [6] gives the following result: if h is a k-automatic function from S_k into a finite set Δ and if A and B are two finite words, then the sequence $(h([0_\bullet AB^n]_k))_{n\geq 0}$ is eventually periodic. This implies that all sequences $(f(a_{ns+j}/k^{ns+j}))_{n\geq 0}$, with

$r \le j < r+s$, are eventually periodic. Consequently, the sequence $(f(a_n/k^n))_{n\ge 1}$ is eventually periodic, concluding the proof. $\square$

Lemma 3.6. *Let $Y \subseteq X \subseteq [0,1]$ be closed sets, a and b be non-negative integers, c and d be non-positive integers, and k and ℓ be two integers larger than* 1. *Let $T : [0,1] \to \mathbb{R}$ and $S : [0,1] \to \mathbb{R}$ be given by $T(x) = kx + c$ and $S(x) = \ell x + d$. Suppose that the following conditions are satisfied:*

(i) *k and ℓ are multiplicatively independent;*
(ii) *$X \cap [b_1, b_2] = Y$ for some $b_1, b_2 \in [0,1]$;*
(iii) *$X = T(X \cap [a/k, a/k + 1/k])$;*
(iv) *$Y = S(Y \cap [b/\ell, (b+1)/\ell])$;*
(v) *$T(\beta) = S(\beta) = \beta$ for some $\beta \in Y$;*
(vi) *β is not the rightmost point of Y, and Y has at least two points in (b_1, b_2).*

Then $X \cap [\beta, 1] = [\beta, 1]$.

Proof. Suppose that Y^c contains an open interval (x, y) with $x, y \in Y$. By (vi), we may assume that $b_1 < x < y < b_2$. By hypothesis (i), for every positive real number ε, the set

$$\mathcal{N} = \{(m,n) \in \mathbb{N}^2 \mid k^m \ < \ \ell^n \ < \ (1+\varepsilon)k^m\} \tag{2}$$

is infinite.

From now on, we fix a positive ε such that

$$\varepsilon < (y-x)/2, \ x > b_1(1+\varepsilon) \text{ and } b_2 > y(1+\varepsilon). \tag{3}$$

Let $(m,n) \in \mathcal{N}$. We consider the set $T^m(S^{-n}(Y))$. By (iv), we have $S^{-n}(Y) \subseteq Y \subseteq X$, while (iii) implies that $T^m(X) \cap [0,1] = X$. Hence we infer from (ii) that

$$T^m(S^{-n}(Y)) \cap [b_1, b_2] \subseteq Y.$$

On the other hand, (ii) and (v) give that $\beta \in [b_1, b_2]$.

Then we have

$$T^m(S^{-n}(z)) = k^m \ell^{-n} z + \beta(1 - k^m \ell^{-n}) \in Y,$$

whenever

$$b_1 \le k^m \ell^{-n} z + \beta(1 - k^m \ell^{-n}) \le b_2.$$

Notice that since $(m,n) \in \mathcal{N}$, (2) and (3) give

$$T^m(S^{-n}(x)) = k^m \ell^{-n} x + \beta(1 - k^m \ell^{-n}) < x + \varepsilon < b < b_2$$

and

$$T^m(S^{-n}(x)) = k^m \ell^{-n} x + \beta(1 - k^m \ell^{-n}) > x/(1+\varepsilon) > b_1.$$

Thus

$$T^m(S^{-n}(x)) \in Y \text{ and } T^m(S^{-n}(x)) < y.$$

Since $(x,y) \subseteq Y^c$, we see that $T^m(S^{-n}(x)) \le x$. Similarly, $T^m(S^{-n}(y)) \in Y$ and

$$T^m(S^{-n}(y)) = k^m \ell^{-n} y + \beta(1 - k^m \ell^{-n}) > x.$$

And again, since $(x,y) \subseteq Y^c$, we obtain $T^m(S^{-n}(y)) > y$. These two inequalities give

$$T^m(S^{-n}(y)) - T^m(S^{-n}(x)) = k^m \ell^{-n}(y - x) \ge y - x.$$

This provides a contradiction, since (2) ensures that $k^m \ell^{-n} < 1$.

It follows that Y^c cannot contain an open interval (x,y) with $x, y \in Y$. Thus Y is either empty or consists of a single closed interval. Since, by (vi), $\beta \in Y$ is not the rightmost point in Y, Y is a closed interval containing $[\beta, \beta+\delta]$ for some positive δ.

We claim that $[\beta, 1] \subseteq X$. Indeed, if $z \in (\beta, 1)$ is not in X, then $(T^{-n}(z))_{n \geq 1}$ is a decreasing sequence converging to β, and thus $T^{-n}(z) \in [\beta, \beta+\delta]$ for every integer n large enough. But, we just obtained that $[\beta, \beta+\delta] \subseteq Y \subseteq X$. This would provide a contradiction since, by (iii), $T^{-1}(X^c) \cap [0,1] \subseteq X^c$. Thus $(\beta, 1) \subseteq X$, and since assumption X is a closed set, we obtain that $[\beta, 1] \subseteq X$. This ends the proof. $\square$

Lemma 3.7. *Let k and ℓ be two multiplicatively independent natural numbers and suppose that $X \subseteq [0,1]$ is a compact k- and ℓ-self-similar set. If X is not a finite union of closed intervals with rational endpoints, then there exists a compact set $\widetilde{X} \subseteq [0,1]$ which is both k- and ℓ-self-similar, and which has a rational wild point with purely periodic base-k and base-ℓ expansions.*

Proof. Since X^c is open, it is a countable disjoint union of open intervals. If the number of open intervals is finite, we obtain that X is a finite union of closed intervals. By Proposition 3.1, this would provide a contradiction with our assumption. So we may assume that we have a countably infinite set of closed intervals. Let

$$\{(\alpha_i, \beta_i) \mid i \in \mathbb{N}\}$$

be an enumeration of these intervals. We have that $\beta_i \in [0,1]$ for infinitely many i, and so by the Bolzano-Weierstrass theorem there exists some number $\beta \in [0,1]$ that is a limit point of the set of β_i. Write

$$\beta \ = \ \sum_{i \geq 1} b_i/k^i,$$

with $0 \leq b_i < k$. Then β is a wild point of X. For every positive integer n, we set

$$c_n := b_1 k^{n-1} + b_2 k^{n-2} + \cdots + b_n.$$

Then

$$c_n/k^n \leq \beta \leq c_n/k^n + 1/k^n.$$

By assumption, the set X is k-self-similar. Thus there are only finitely many distinct sets of the form

$$\{k^n x - c_n \mid x \in [c_n/k^n, (c_n+1)/k^n] \cap X\},$$

and hence there exist distinct positive integers i and j with $j > i$ such that

$$\begin{aligned} X_0 &:= \{k^i x - c_i \mid x \in [c_i/k^i, (c_i+1)/k^i] \cap X\} \\ &= \{k^j x - c_j \mid x \in [c_j/k^j, (c_j+1)/k^j] \cap X\}. \end{aligned}$$

Note that the definition of X_0 implies that X_0 is also both k- and ℓ-self-similar, and that $\beta' := k^i\beta - c_i$ is a wild point of X_0.

Let $m = j - i$. Then

$$\begin{aligned} X_0 &= \{k^j x - c_j \mid x \in [c_j/k^j, (c_j+1)/k^j] \cap X\} \\ &= \{k^m(k^i x - c_i) - c_j + k^m c_i \mid x \in [c_j/k^j, (c_j+1)/k^j] \cap X\} \\ &= \{k^m y - c_j + k^m c_i \mid y \in [c_j/k^m - c_i, c_j/k^m - c_i + 1/k^m] \cap (k^i X - c_i)\}. \end{aligned}$$

By definition, $X_0 = (k^i X - c_i) \cap [0,1]$, and so

$$X_0 \ = \ \{k^m x + k^m c_i - c_j \mid x \in [c_j/k^m - c_i, c_j/k^m - c_i + 1/k^m] \cap X_0\}.$$

Thus we have a map $U : [0,1] \to [0,1]$ given by

$$U(x) = (x + c_j - k^m c_i)/k^m,$$

which satisfies $U(X_0) \subseteq X_0$ and $U(X_0^c) \subseteq X_0^c$. Let $c = c_j k^{-m} - c_i$. By Lemma 3.4, $U(\beta'), U^2(\beta'), \ldots$ are all wild points of X_0. Moreover, this sequence is bounded and monotonic, and hence it converges to the unique fixed point of U; namely, the point

$$\gamma := k^m c/(k^m - 1). \tag{4}$$

We thus infer from Lemma 3.4 that $\gamma = k^m c/(k^m - 1)$ is also a wild point of X_0.

Since γ is rational, we can write

$$\gamma \ = \ a/\ell^p + a'/\ell^p(\ell^p - 1)$$

for some natural numbers a and a', and some positive integer p such that $a' < \ell^p - 1$. For every positive integer n, set

$$a_n := \lfloor \ell^{pn}\gamma \rfloor.$$

We are now going to use the fact that the set X_0 is also ℓ-self-similar. Hence there are only finitely many distinct sets of the form

$$\{\ell^{np}x - a_n \mid x \in [a_n/\ell^{pn}, (a_n+1)/\ell^{pn}] \cap X_0\}.$$

Let $Y_1, \ldots, Y_q$ denote the distinct sets of this form. By Lemma 3.5, if $f(n)$ denotes the index such that

$$Y_{f(n)} = \{\ell^{np}x - a_n \mid x \in [a_n/\ell^{pn}, (a_n+1)/\ell^{pn}] \cap X_0\},$$

then the sequence $(f(n))_{n\geq 1}$ is eventually periodic. Hence there exist some index i and some positive integer d such that

$$Y_i = \{\ell^{ndp}x - a_{ndp} \mid x \in [a_{nd}/\ell^{ndp}, (a_{nd}+1)/\ell^{ndp}] \cap X_0\} \tag{5}$$

for all $n \geq 1$. Since X_0 is both k- and l-self-similar, we obtain that Y_i is also both k- and l-self-similar. Moreover, since γ is a wild point of X_0, we also get that $\gamma' := \ell^{dp}\gamma - a_{dp}$ is a wild point of Y_i. Furthermore, we have

$$\gamma' = a'/(\ell^p - 1).$$

To finish the proof, note that

$$\gamma = a/\ell^p + a'/\ell^p(\ell^p - 1)$$

has the property that $(k^m - 1)\gamma \in \mathbb{Z}$. Thus $(k^m - 1)\ell^p\gamma = (k^m - 1)a + (k^m - 1)a'/(\ell^p - 1)$ is an integer. Consequently, $(k^m - 1)a'/(\ell^p - 1)$ is an integer. But this means that γ' has both a purely periodic base-k and base-ℓ expansion, concluding the proof. $\square$

We are now ready to prove our main result.

Proof of Theorem 1.4. Let k and l be two multiplicatively independent integers, and let X be a closed subset of $[0,1]$ which is both k- and l-self-similar. We are going to argue by contradiction.

Let us assume that X is not a finite union of closed intervals with rational endpoints. By Lemma 3.7, there exists a compact set $\widetilde{X} \subseteq [0,1]$ which is both k- and ℓ-self-similar and which has a wild point, say β, whose base-k and base-ℓ expansions are purely periodic. By replacing k and ℓ with an appropriate power, we may assume that this period is 1 for each base. It is also no loss of generality to assume that β is a right wild point of $\widetilde{X}$. Indeed, if $\widetilde{X}$ is k-self-similar, then the set $1-\widetilde{X} := \{x \in [0,1] \mid 1-x \in \widetilde{X}\}$ is also k-self-similar, so that we could if necessary replace $\widetilde{X}$ by $1-\widetilde{X}$.

For every integer $n \geq 1$ there exist non-negative integers a_n and b_n such that

$$a_n/k^n \leq \beta \leq a_n/k^n + 1/k^n \tag{6}$$

and

$$b_n/\ell^n \leq \beta \leq b_n/\ell^n + 1/\ell^n. \tag{7}$$

We first use the k-self-similarity of $\widetilde{X}$. Since $\widetilde{X}$ is k-self-similar, there are only finitely many distinct sets of the form

$$\{k^n x - a_n \mid x \in [a_n/k^n, (a_n+1)/k^n] \cap \widetilde{X}\},$$

and hence there exist distinct positive integers i and j, with $j > i$, and such that

$$\begin{aligned} X_0 &:= \{k^i x - a_i \mid x \in [a_i/k^i, (a_i+1)/k^i] \cap \widetilde{X}\} \\ &= \{k^j x - a_j \mid x \in [a_j/k^j, (a_j+1)/k^j] \cap \widetilde{X}\}. \end{aligned}$$

It thus follows from this definition that X_0 is both k- and ℓ-self-similar. Furthermore, since β has a purely periodic base-k expansion, we also have that $\beta = k^i\beta - a_i$ is a wild point of X_0.

Set $m := j - i$. Then

$$\begin{aligned} X_0 &= \{k^j x - a_j \mid x \in [a_j/k^j, (a_j+1)/k^j] \cap \widetilde{X}\} \\ &= \{k^m(k^i x - a_i) - a_j + k^m a_i \mid x \in [a_j/k^j, (a_j+1)/k^j] \cap \widetilde{X}\} \\ &= \{k^m y - a_j + k^m c_i \mid y \in [a_j/k^m - c_i, a_j/k^m - a_i + 1/k^m] \cap (k^i X - a_i)\}. \end{aligned}$$

By assumption, $X_0 = (k^i\widetilde{X} - a_i) \cap [0,1]$, and so

$$X_0 = \{k^m x + k^m a_i - a_j \mid x \in [a_j/k^m - a_i, a_j/k^m - a_i + 1/k^m] \cap X_0\}. \tag{8}$$

Set $c := a_j/k^m - a_i$, and let $T : [0,1] \to \mathbb{R}$ be the map defined by

$$T(x) = k^m x - k^m c.$$

Then we infer from equality (8) that

$$X_0 = T([c, c+1/k^m] \cap X_0). \tag{9}$$

Furthermore, since β has a purely periodic expansion of period 1 in base k, we infer from (6) that

$$T(\beta) = \beta. \tag{10}$$

We are now going to use the ℓ-self-similarity of X_0. Since X_0 is ℓ-self-similar, there are only finitely many distinct sets of the form

$$\{\ell^n x - b_n \mid x \in [b_n/\ell^n, (b_n+1)/\ell^n] \cap X_0\}.$$

Let $Y_1, \ldots, Y_q$ denote the distinct sets of this form. Let $f(n)$ denote the index such that

$$Y_{f(n)} = \{\ell^n x - b_n \mid x \in [b_n/\ell^n, (b_n+1)/\ell^n] \cap X_0\}.$$

Then it follows from Lemma 3.5 that the sequence $(f(n))_{n\geq 1}$ is eventually periodic. Hence there exist some index i and some positive integer d such that

$$Y_i = \{\ell^{dn} x - b_{dn} \mid x \in [b_{dn}/\ell^{dn}, (b_{dn}+1)/\ell^{dn}] \cap X_0\} \tag{11}$$

for all $n \geq 1$. To complete the proof, we define a subset Y of X_0 by

$$Y := X_0 \cap [b_d/\ell^d, (b_d+1)/\ell^d]. \tag{12}$$

Let $S : [0,1] \to \mathbb{R}$ be the map defined by

$$S(x) = \ell^d x - b_d.$$

Since β has a purely periodic base-ℓ expansion of period 1, we see that

$$S(\beta) = \beta. \tag{13}$$

We also claim that

$$S(Y \cap [b_{2d}/\ell^{2d}, (b_{2d}+1)/\ell^{2d}]) = Y. \tag{14}$$

To see this, note that by (11) we have $Y_i = S(X_0 \cap [b_d/\ell^d, (b_d+1)/\ell^d]) = S(Y)$ and also

$$Y_i = S^2(X_0 \cap [b_{2d}/\ell^{2d}, (b_{2d}+1)/\ell^{2d}]) = S^2(Y \cap [b_{2d}/\ell^{2d}, (b_{2d}+1)/\ell^{2d}]),$$

since $[b_{2d}/\ell^{2d}, (b_{2d}+1)/\ell^{2d}] \subset [b_d/\ell^d, (b_d+1)/\ell^d]$. Thus

$$S(Y) = S^2(Y \cap [b_{2d}/\ell^{2d}, (b_{2d}+1)/\ell^{2d}]),$$

which implies equality (14) since S is an injective map.

Since β is a right wild point of X_0, we see that it is not the rightmost point in Y. We are now ready to apply Lemma 3.6. Indeed, we infer from equalities (9), (10), (12), (13), (14) and from Lemma 3.6 that $X_0 \cap [\beta, 1] = [\beta, 1]$. This provides a contradiction since β is a right wild point of X_0.

We thus have proved that X is a finite union of closed intervals with rational endpoints. In view of Propositions 3.1 and 3.2, this ends the proof of Theorem 1.4. □

4. Entropy and Hausdorff dimension

In this section, we discuss two notions that can be naturally attached to automatic fractals.

An important notion in the study of fractals is the *Hausdorff dimension*. We do not recall the definition of Hausdorff definition and instead refer the reader to Falconer [12] or to Rogers [20] for an introduction to this topic. One fact about most fractal objects is that their Hausdorff dimension (or the Hausdorff dimension of their boundary) is not a natural number. For instance, the Hausdorff dimension of the triadic Cantor set is equal to $\log 2/\log 3$, and we, respectively, get the values $\log 8/\log 3$, $\log 3/\log 2$ and $\log 20/\log 3$ for the Hausdorff dimension of the Sierpiński carpet, Pascal's triangle modulo 2, and the Menger sponge.

Mauldin and Williams [18] studied a large family of fractal sets of $\mathbb{R}^n$, which they called geometric graph directed constructions. These fractals are constructed by means of a directed labelled graph. Among other results, Mauldin and Williams showed that the Hausdorff dimension of such a set can be computed by working on the so-called weighted incidence matrix associated with the graph.

A directed labeled graph is associated to a finite automaton in a natural way. This can be used to show that every automatic fractal can be obtained as a geometric graph directed construction. For an automatic fractal X, it should thus be possible to use the approach of Mauldin and Williams to compute $\mathcal{H}(X)$, the Hausdorff dimension of X.

Two fundamental notions in the theory of formal languages are the *subword complexity* and the *entropy* of factorial languages. Given a finite set $\mathcal{A}$, a language $\mathcal{L}$ over $\mathcal{A}$ is just a subset of $\mathcal{A}^*$. The language $\mathcal{L}$ is called *factorial* if for every word W over $\mathcal{L}$, every subword of W also belongs to $\mathcal{L}$. Then the complexity function of a factorial language $\mathcal{L}$ is defined as the function that maps every integer n to the integer

$$p(\mathcal{L}, n) := \#(\mathcal{L} \cap \mathcal{A}^n),$$

and the entropy of the language $\mathcal{L}$ is defined by

$$h(\mathcal{L}) := \lim_{n\to\infty} \frac{1}{n} \log p(\mathcal{L}, n).$$

The latter notion is well-defined; indeed, for a factorial language the limit above always exists because of the trivial inequality $p(\mathcal{L}, n+m) \le p(\mathcal{L}, n) \times p(\mathcal{L}, m)$. With this definition we always have

$$0 \le h(\mathcal{L}) \le \log(\#\mathcal{A}).$$

We now explain how to associate a factorial language with an automatic fractal. Let X be a k-automatic fractal of $\mathbb{R}^d$. Let $\underline{x} := (x_1, \ldots, x_d)$ be an element of $X \cap S_k^d$. Then expanding each coordinate in base k, we obtain the existence of finite words $W_1, \ldots, W_d \in \Sigma_k^*$ such that

$$x_i = [_\bullet W_i]_k$$

for every integer $1 \le j \le d$. Set

$$W_i 0^\infty := w_1^{(i)} w_2^{(i)} \cdots .$$

Thus

$$\underline{x} = ([_\bullet W_1 0^\infty]_k, [_\bullet W_2 0^\infty]_k, \ldots, [_\bullet W_d 0^\infty]_k),$$

and we can associate with $\underline{x}$ an element $\mathbf{w}(\underline{x})$ of $(\Sigma_k^d)^{\mathbb{N}}$ defined by

$$\mathbf{w}(\underline{x}) := \begin{pmatrix} w_1^{(1)} \\ \vdots \\ w_1^{(d)} \end{pmatrix} \begin{pmatrix} w_2^{(1)} \\ \vdots \\ w_2^{(d)} \end{pmatrix} \cdots \begin{pmatrix} w_n^{(1)} \\ \vdots \\ w_n^{(d)} \end{pmatrix} \cdots .$$

We then consider the language $\mathcal{L}(X)$ formed by all finite words in $(\Sigma_k^d)^*$ having at least one occurrence in some $\mathbf{w}(\underline{x})$, with $\underline{x} \in X \cap S_k^d$. By definition, $\mathcal{L}(X)$ is a factorial language. We thus define the entropy of the automatic fractal X to be

$$h(X) := \frac{1}{\log k}\, h(\mathcal{L}(X)).$$

Recall that a k-automatic fractal is also k^n-automatic for every positive integer n. With the normalization above, the entropy of a k-automatic fractal remains unchanged when viewing it as a k^n-automatic fractal.

Since X is a k-automatic fractal, it can be shown that the language $\mathcal{L}(X)$ is recognized by a finite automaton (see Sakarovitch [21]). Then Kleene's theorem implies that $\mathcal{L}(X)$ is a rational language, and by a result of Schützenberger, we get that the formal power series

$$\sum_{n\ge 1}^{+\infty} p(\mathcal{L}(X), n) X^n$$

is a rational function; that is, it belongs to $\mathbb{Q}(X)$ (see Eilenberg [11]). In this case, the sequence $(p(\mathcal{L}(X), n))_{n\geq 1}$ satisfies a linear recurrence, and this can be used to compute the entropy of $\mathcal{L}(X)$.

The Hausdorff dimension of an automatic fractal and its entropy appear to be strongly connected. Let us give a few examples.

With the previous definition, we easily obtain that the language associated with the triadic Cantor set is

$$\mathcal{L}(C) = \{a_1 a_2 \cdots a_k \mid a_i \in \{0, 2\},\ 1 \leq i \leq k\}.$$

Consequently, $p(\mathcal{L}(C), n) = 2^n$, and thus

$$h(C) = \log 2/\log 3 = \mathcal{H}(C).$$

Let S denote the Sierpiński carpet defined in Example 2.7. We easily obtain that

$$\mathcal{L}(S) = \left\{ \begin{pmatrix} a_1 \\ b_1 \end{pmatrix} \begin{pmatrix} a_2 \\ b_2 \end{pmatrix} \cdots \begin{pmatrix} a_n \\ b_n \end{pmatrix} \in (\Sigma_3^2)^* \mid \begin{pmatrix} a_i \\ b_i \end{pmatrix} \neq \begin{pmatrix} 1 \\ 1 \end{pmatrix},\ 1 \leq i \leq n \right\}.$$

Thus we get that $p(\mathcal{L}(S), n) = 8^n$ for every positive integer n, and consequently

$$h(S) = \log 8/\log 3 = \mathcal{H}(S).$$

A similar computation with the Menger sponge M leads to equalities

$$h(M) = \log 20/\log 3 = \mathcal{H}(M).$$

It would be interesting to determine the exact link between the entropy and the Hausdorff dimension of automatic fractals. In particular, one may ask whether the equality

$$h(X) = \mathcal{H}(X)$$

holds for every automatic fractal X.

5. Comments

The interplay between fractal sets and finite automata has quite a long history. It is not our purpose to give a survey of such studies, and so we give only a few references.

Our definition of a k-automatic fractal uses two steps. First, we made the choice to consider finite-state automata as devices that only take finite words as input. In this way, we associate a subset of the non-negative k-adic rationals with a finite automaton A that takes words over the alphabet Σ'_k as input. Then we obtain a k-automatic fractal X by taking the closure of this set. Another equivalent formulation can actually be given in terms of Büchi automata. More precisely, we would obtain the same automatic fractal X by considering the set of real numbers in $[0, 1]$ whose base-k expansion has the property that either it is finite and accepted by A or it is infinite and infinitely many of its prefixes are accepted by A.

Hartmanis and Stearns [16] were probably the first to use finite automata to describe some fractal subsets of $[0, 1]$. Their approach, though different from ours, is in the same spirit as the one described just above in terms of Büchi automata. Note also that these authors were only concerned with one-dimensional sets. With a finite k-automaton A, they associate the set of real numbers such that all prefixes in their base-k expansion are recognized by A. In many cases, the set they get using this construction is the same as the one obtained by using the process described in Section 2. However, this is not always the case. For instance, starting with the Thue–Morse automaton, they get the set $\{1\}$, while our construction gives the whole interval $[0, 1]$.

In contrast, most constructions of fractals involving finite automata follow a rather different route. Starting with a classical automatic function f of $\mathbb{N}^d$, a sequence of arrays (or matrices) corresponding to compact sets of $\mathbb{R}^d$ is naturally associated with f. After some kind of renormalization, this leads to a sequence of compact sets of $[0, 1]^d$. Then fractals are obtained as those sets that are a limit point of this sequence of compact sets with respect to the Hausdorff metric. Many authors already used this principle on specific examples such as the Sierpiński carpet or Pascal's triangle [24, 5, 4] (see also Allouche and Shallit [6, Chapter 14]). More recent and general accounts were given by von Haeseler *et al.* [14, 15] and by Barbé and von Haeseler [7]. In particular, Barbé and von Haeseler [7] include a systematic study of automatic fractals. By the process just described, these authors obtain a family of fractals which is essentially the same as the family of automatic fractals we defined in the present paper. In particular, our results would also apply in their framework.

Note that another geometric context in which a Cobham-type phenomenon occurs is given by stretching factors of self-similar tilings, as recently described by Cortez and Durand [10].

After writing a first draft of this article, we learned about recent work of Boigelot and Brusten [8]. Although they are not motivated in any way by fractals and their results are written in terms of logic, there is no doubt an intimate connection with the present work.

We end this paper with two questions.

Problem 5.1. Let X be an automatic fractal of $\mathbb{R}^d$ whose Hausdorff dimension is d. Does the set X always contain a non-empty open set?

The following problem is motivated by classical questions in number theory concerning the expansion of algebraic irrational numbers in integer bases.

Problem 5.2. Is it true that an automatic fractal of $\mathbb{R}$ contains an algebraic irrational number only if it contains an open interval? Note that it would already be interesting to find a one-dimensional automatic fractal X such that $0 < \mathcal{H}(X) < 1$, and for which one can prove that it contains no irrational algebraic numbers.

In the case where X is the triadic Cantor set, the latter question corresponds to a famous problem addressed by Mahler [19]. Roughly, we could say that if an automatic set X of $\mathbb{R}$ has Hausdorff dimension less than one, then all elements in X should have serious restrictions regarding patterns that can occur in their base-k expansion. On the other hand, it is expected that algebraic irrational numbers contain every possible finite sequence of digits in their base-k expansion (see, for instance, Adamczewski and Bugeaud [2]).

Acknowledgments

We thank the referee for making many helpful and insightful comments which have greatly improved this paper.

References

[1] B. Adamczewski and Y. Bugeaud, Real and p-adic expansions involving symmetric patterns, *Int. Math. Res. Not.*, Volume 2006 (2006), Article ID 75968, 17 pages. MR2250005 (2007e:11086)

[2] B. Adamczewski and Y. Bugeaud, On the complexity of algebraic numbers I. Expansions in integer bases, *Annals of Math.* **165** (2007), 547–565. MR2299740 (2008a:11130)

[3] P. S. Addison, *Fractals and chaos. An illustrated course.* Institute of Physics Publishing, Bristol, 1997. MR1483313 (98k:58152)

[4] J.-P. Allouche, F. von Haeseler, H.-O. Peitgen, A. Petersen and G. Skordev, Automaticity of double sequences generated by one-dimensional linear cellular automata, *Theoret. Comput. Sci.* **188** (1997), 195–209. MR1479629 (98j:68116)

[5] J.-P. Allouche, F. von Haeseler, H.-O. Peitgen and G. Skordev, Linear cellular automata, finite automata and Pascal's triangle, *Discrete Appl. Math.* **66** (1996), 1–22. MR1387674 (97b:11033)

[6] J.-P. Allouche and J. Shallit, *Automatic Sequences: Theory, Applications, Generalizations*, Cambridge University Press, Cambridge, 2003. MR1997038 (2004k:11028)

[7] A. Barbé and F. von Haeseler, Limit sets of automatic sequences. *Adv. Math.* **175** (2003), 169–196. MR1972631 (2004b:11026)

[8] B. Boigelot and J. Brusten, A generalization of Cobham's theorem to automata over real numbers, *Theoret. Comput. Sci.* **410** (2009), 1694–1703. MR2508527 (2010f:68133)

[9] A. Cobham, On the base-dependence of sets of numbers recognizable by finite automata, *Math. Systems Theory* **3** (1969), 186–192. MR0250789 (40:4021)

[10] M. I. Cortez and F. Durand, Self-similar tiling systems, topological factors and stretching factors, *Discrete Comput. Geom.* **40** (2008), 622–640. MR2453331 (2009i:52013)

[11] S. Eilenberg, *Automata, languages, and machines. Vol. A*, Pure and Applied Mathematics, Vol. 58, Academic Press, New York, 1974. MR0530382 (58:26604a)

[12] K. Falconer, *Fractal geometry. Mathematical foundations and applications*, John Wiley & Sons, Ltd., Chichester, 1990. MR1102677 (92j:28008)

[13] H. Furstenberg, Disjointness in ergodic theory, minimal sets, and a problem in Diophantine approximation, *Math. Systems Theory* **1** (1967), 1–49. MR0213508 (35:4369)

[14] F. von Haeseler, H.-O. Peitgen and G. Skordev, Self-similar structure of rescaled evolution sets of cellular automata. I, *Int. J. Bifurcation Chaos Appl. Sci. Eng.* **11** (2001), 913–926. MR1839903 (2003e:37013)

[15] F. von Haeseler, H.-O. Peitgen and G. Skordev, Self-similar structure of rescaled evolution sets of cellular automata. II, *Int. J. Bifurcation Chaos Appl. Sci. Eng.* **11** (2001), 927–941. MR1839904 (2003e:37014)

[16] J. Hartmanis and R. E. Stearns, Sets of numbers defined by finite automata, *Amer. Math. Monthly* **74** (1967), 539–542. MR0211817 (35:2694)

[17] K. Kedlaya, Finite automata and algebraic extensions of function fields, *J. Théor. Nombres Bordeaux* **18** (2006), 379–420. MR2289431 (2007m:11044)

[18] D. R. Mauldin and S. C. Williams, Hausdorff dimension in graph directed constructions, *Trans. Amer. Math. Soc.* **309** (1988), no. 2, 811–829. MR961615 (89i:28003)

[19] K. Mahler, Some suggestions for further research, *Bull. Austral. Math. Soc.* **29** (1984), 101–108. MR732177 (85e:11002)

[20] C. A. Rogers, *Hausdorff measures*, Cambridge University Press, New York, 1970. MR0281862 (43:7576)

[21] J. Sakarovitch, Éléments de théorie des automates, Vuibert, Paris, 2003 (English translation: Elements of Automata Theory, Cambridge University Press, 2009).

[22] O. Salon, Suites automatiques à multi–indices et algébricité, *C. R. Acad. Sci. Paris* **305** (1987), 501–504. MR916320 (88k:11094)

[23] A. L. Semenov, The Presburger nature of predicates that are regular in two number systems, *Siberian Mathematical Journal* **18** (1977), 289–299. MR0450050 (56:8349)

[24] J. Shallit and J. Stolfi, Two methods for generating fractals, *Comput. & Graphics* **13** (1989), 185–191.

[25] M. Waldschmidt, *Diophantine approximation on linear algebraic groups*, Springer-Verlag, 2000. MR1756786 (2001c:11075)

[26] M. Waldschmidt, *Un demi-siècle de transcendance*. In: Development of mathematics 1950–2000, pp. 1121–1186, Birkhäuser, Basel, 2000. MR1796871 (2001k:11130)

[27] L. Wegner, Problem P12: Is www^R cube-free? *Bull. Eur. Assoc. Theor. Comput. Sci.*, No. 18 (October 1982), 120.

CNRS, Université de Lyon, Université Lyon 1, Institut Camille Jordan, 43 boulevard du 11 novembre 1918, 69622 Villeurbanne Cedex, France

E-mail address: Boris.Adamczewski@math.univ-lyon1.fr

Department of Mathematics, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia, Canada V5A 1S6

E-mail address: jpb@math.sfu.ca