BI-LIPSHITZ EMBEDDING OF ULTRAMETRIC CANTOR SETS INTO EUCLIDEAN SPACES

JEAN V. BELLISSARD, ANTOINE JULIEN

ABSTRACT. Let (C, d) be an ultrametric Cantor set. Then it admits an isometric embedding into an infinite dimensional Euclidean space [30]. Associated with it is a weighted rooted tree, the reduced Michon graph $\mathcal{T}$ [28]. It will be called f-embeddable if there is a bi-Lipshitz map from (C, d) into a finite dimensional Euclidean space. The main result establishes that (C, d) is f-embeddable if and only if it can be represented by a weighted Michon tree such that (i) the number of children per vertex is uniformly bounded, (ii) if κ denotes the weight, there are constants $c > 0$ and $0 < \delta < 1$ such that $\kappa(v)/\kappa(u) \leq c\,\delta^{d(u,v)}$ where v is a descendant of u and where $d(u, v)$ denotes the graph distance between the vertices u, v. Several examples are provided: (a) the tiling space of a linear repetitive sequence is f-embeddable, (b) the tiling space of a Sturmian sequence is f-embeddable if and only if the irrational number characterizing it has bounded type, (c) the boundary of a Galton-Watson random trees with more than one child per vertex is almost surely not f-embeddable.

1. INTRODUCTION

Embedding isometrically compact metric spaces into a Euclidean space is an old problem [17, 6, 33, 34] that has been investigated by several communities, in Mathematics and in Computer Science as well. There are elementary examples of finite metric space for which such an embedding is impossible: the simplest example is the space $K_{1,3} = \{0, 1, 2, 3\}$ (see Fig. 1 and [21]) with metric $d(0, i) = 1$ for $i = 1, 2, 3$ and $d(i, j) = 2$ if $i \neq j \in X \setminus \{0\}$. This is the path metric on the graph with three vertices and the three internal edges. Then there is no isometric embedding of (X, d) in any Euclidean space. More generally,

Theorem 1 (Schoenberg [32])**.** *If (X, d) is a finite metric space, let $x \in X$ be a point and let $A(x)$ be the matrix $A(x)$ indexed by $X \setminus \{x\}$ with elements $A(x)_{y,z} = (d(x, y)^2 + d(x, z)^2 - d(y, z)^2)/2$. Then (X, d) admits an isometric embedding in $\mathbb{R}^n$ if and only if*
  (i) $A(x)$ is a positive matrix and
  (ii) the rank of $A(x)$ is less than or equal to n.

Testing whether $A(x)$ is positive can be done through the Cholesky algorithm, which provides also a computation of the rank. This method is amenable to numerical calculation to analyze data sets (see [27, 8] for instance).

The simplest type of infinite metric spaces for which the isometrical embedding problem can be extended are Cantor sets. A Cantor set is a compact Hausdorff space that is completely disconnected and has no isolated point. An old Theorem of Brouwer [7] claims that all Cantor sets are homeomorphic to the set $\{0, 1\}^{\mathbb{N}}$ of infinite sequences of 0's and 1's, endowed with the
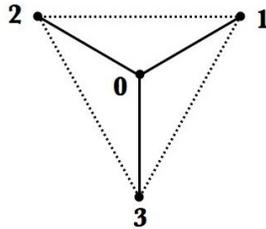
FIGURE 1. The space $K_{1,3}$ with $d(0,i) = 1$, $d(i,j) = 2$ is not embeddable

product topology. While there is only one Cantor set, as a topological space, there are many metrics on it describing the topology. The subclass of ultrametrics turns out to be as natural for Cantor sets as the Euclidean metric on $\mathbb{R}^n$. This was investigated in the eighties by Michon [28] who described all possible ultrametrics through the concept of weighted rooted tree. In a recent work [30], J. Pearson and the first author, using this formalism, proved the following embedding result

Theorem 2 (see [30]). *Any ultrametric Cantor set can be embedded isometrically into an infinite dimensional Euclidean space with a countable basis.*

The algorithm behind this result also applies trivially to finite ultrametric spaces. Actually, in a finite space X, all metric are equivalent, so that replacing the metric by the smallest ultrametric which dominates it, allows to use the method of [30] to find a quasi-isometric embedding in a finite dimensional Euclidean space. The method however gives a dimension n which is at least the number of points of the space minus one. On the other hand, given any finite set F on the real line with the same cardinality as X and given any one-to-one map from X to F, the metric on X induced by the Euclidean metric on F is equivalent to the original one, so that quasi isometric embedding of X in $\mathbb{R}$ is always possible. As it turns out this is a subclass of sets for which the main result (Theorem 3) of the present work applies.

One of the questions suggested by Theorem 2 is whether the dimension of the embedding space can be made finite. There is an obvious obstacle to such an embedding: since the Cantor set is ultrametric, it cannot be embedded isometrically into $\mathbb{R}^n$, because the Euclidean metric is not an ultrametric. However, it is sufficient that the two metrics be equivalent. Again there is an obvious obstacle, namely if such an embedding exists, then the Hausdorff dimension of (C,d) and of its image $\phi(C)$ are equal and therefore $\dim_H(C,d) \leq n$. So the question has to be rephrased as follows

Definition 1. *An ultrametric Cantor set (C,d) will be called f-embeddable if there is $n \in \mathbb{N}$, a positive real number c and a map $\phi : C \to \mathbb{R}^n$ such that*

$$\frac{1}{c}\, d(x,y) \leq \|\phi(x) - \phi(y)\| \leq c\, d(x,y)\,, \qquad\qquad \forall x,y \in C\,.$$

Such a map will be called a quasi-isometry *or a* bi-Lipshitz *map, or, for short, in this paper, an f-embedding. Then n will be called the* dimension *of the embedding.*

2. The Main Result

A graph $\mathcal{G}$ is a family $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$, called the set of vertices, and $\mathcal{E}$, called the set of edges, are countable sets. They occur with two maps $(r, s) : \mathcal{E} \to \mathcal{V}$, everywhere defined, called the range and the source. A finite path in $\mathcal{G}$ is a sequence $\gamma = (e_1, \cdots, e_n)$ of edges such that $r(e_j) = s(e_{j+1})$. The integer $n = |\gamma|$ is called the length of γ. The maps r, s extend to finite path by $r(\gamma) = r(e_n)$ and $s(\gamma) = s(e_1)$. If $x = s(\gamma)$ and $y = r(\gamma)$ then γ connects x to y and this will be denoted by $\gamma : x \to y$. If $v = r(e_i) = s(e_{i+1})$, then γ is said to pass through v. A path is a *loop* if $r(\gamma) = s(\gamma)$. The relation "*x is connected to y*", defined by "there is a path γ such that either $\gamma : x \to y$ or $\gamma : y \to x$", is an equivalence. The equivalence classes of $\mathcal{G}$ are called the connected components. If there is only one connected component, $\mathcal{G}$ is called *connected*. A path γ is a *cycle* if it is a non-trivial path $\gamma : x \to x$. In particular, a loop is a cycle.

A tree $\mathcal{T}$ is a connected graph with no cycle. A rooted tree is a tree with a pointed vertex $\emptyset$ called the root. In such a case, given a vertex $v \in \mathcal{V}$ there is a unique path connecting the root to v. Then $|v|$ will denote the length of this path and will be called the *generation* of v. This induces an order between the vertices as follows: two vertices v, w satisfies $w \preceq v$ whenever the unique path from $\emptyset$ to w is passing through v. Then v is called an *ancestor* of w, while w is called a *descendant* of v. In such a case, v is the father of w (or, equivalently, w is a child of v) whenever the unique path from v to w contains only one edge. A vertex is *dangling* whenever it has no descendant. A vertex is *branching* if it has at least two children. Given a rooted tree $\mathcal{T} = (\mathcal{V}, \mathcal{E}, \emptyset)$, let $\mathcal{V}' \subset \mathcal{V}$ be the set of branching vertices, which have infinitely many branching descendants. If $w \in \mathcal{V}'$, define its predecessor v to be the closest ancestor in $\mathcal{V}'$. This leads to the *reduced* tree $\mathcal{T}' = (\mathcal{V}', \mathcal{E}', \emptyset')$ where $e \in \mathcal{E}'$ is a pair $(v, w) \in \mathcal{V}' \times \mathcal{V}'$, such that v is the predecessor of w. In such a case $s(v, w) = v$ while $r(v, w) = w$. The new root $\emptyset'$ is either $\emptyset$ if the latter belongs to $\mathcal{V}'$ or its closest descendant belonging to $\mathcal{V}'$. All the trees considered here will be reduced, unless specified otherwise. A rooted tree is called *Cantorian* whenever each vertex has only a finite number of children and each vertex admits a descendant having more than one child.

Let $\mathcal{T}$ be a rooted tree. Its boundary $\partial \mathcal{T}$ is the set of maximal paths starting at the root. If $\mathcal{T}$ has dangling vertices, the latter are representatives of a point in $\partial \mathcal{T}$. So $\partial \mathcal{T}$ is the union of the dangling vertices and of the set of infinite maximal paths. Given a vertex v, $[v]$ will denote the set of paths in $\partial \mathcal{T}$ passing through v. The family $\{[v] \, ; \, v \in \mathcal{V}\}$ defines a basis for a topology on $\partial \mathcal{T}$ which makes this space completely discontinuous. In particular $\partial \mathcal{T}$ is a Cantor set if and only if $\mathcal{T}$ is Cantorian [30]. Conversely any Cantor set can be described through a rooted tree which is not necessarily unique. If $x, y \in \partial \mathcal{T}$, then $x \wedge y$ will denote the youngest common ancestor of x and y. A weight of the rooted tree $\mathcal{T}$ is a map $\kappa : \mathcal{V} \to (0, \infty)$ such that (i) if $w \preceq v$ then $\kappa(w) \leq \kappa(v)$ and (ii) $\lim_{|v| \to \infty} \kappa(v) = 0$. A weight κ leads to an ultrametric d_κ on $\partial \mathcal{T}$ by setting [28, 30]

$$d_\kappa(x, y) = \kappa(x \wedge y), \qquad\qquad x, y \in \partial \mathcal{T}.$$

Conversely the Michon theorem [28] establishes that any ultrametric Cantor set is isometric to the boundary of a reduced rooted Cantorian tree endowed with a weight. Such a tree will be called a *Michon tree*. The main result can be now expressed more precisely

Theorem 3 (Main Result)**.** *An ultrametric Cantor set (C, d) is f-embeddable if and only if it can be represented by a Michon tree $\mathcal{T}$ with weight κ satisfying the following two conditions*
   (i) the number of children of a vertex is uniformly bounded,

(ii) there are constant $c > 0$ and $0 < \delta < 1$ such that

$$\frac{\kappa(w)}{\kappa(v)} \leq c\,\delta^{|w|-|v|}\,, \qquad \forall w \preceq v \in \mathcal{V}\,.$$

The proof that these conditions are necessary is new, to the best of the authors knowledge. However the proof that they are sufficient is deeply inspired by a recent paper by J. Savinien and the second author [22], in which a similar result was obtained for what the authors are calling *self similar Cantor sets*. Examples of such Cantor sets can be obtained as tiling spaces of substitution tilings [31]. Thanks to the existence of a substitution, the conditions (i) and (ii) are easy to implement and the proof reduces to show the existence of an f-embedding. This proof is actually transferable to the present situation and will show that the two conditions (i) and (ii) are sufficient. In [22] though, the authors have an additional result, namely the dimension of the embedding can be chosen to be the smallest integer larger than the Hausdorff dimension of (C, d). Such a result cannot be expected to be true for a general ultrametric Cantor sets. However, the following result will be shown in Section 5

Proposition 1. *If (C, d) is an ultrametric Cantor set for which the conditions (i) & (ii) of Theorem 3 hold, then*

$$\dim_H(C, d) \leq \frac{\ln M}{|\ln \delta|}$$

where M is the maximum number of children per vertex. Moreover the smallest dimension of an f-embedding is at most equal to $\ln(Mc + 1)/|\ln \delta|$.

Several consequences of the Main Result will be proved at the end of this work. Given a doubly infinite sequence $\xi = (u_n)_{n\in\mathbb{N}}$, of letters taken from a finite alphabet $u_n \in A$, the tiling space of ξ is the set of all doubly infinite sequences sharing the same dictionary of finite words. It is customary to define a metric on the tiling space, called the *combinatorial metric*, as follows: if $\xi = (u_n)_{n\in\mathbb{N}}, \eta = (v_n)_{n\in\mathbb{N}}$ are two sequences their distance is $d(\xi, \eta) = (N + 1)^{-1}$, where N is the largest natural integer $N \geq 0$ such that $u_n = v_n$ for $|n| \leq N$.

The first consequence of the Main Result concerns the tiling space of a *linearly repetitive* sequence. A sequence $x = (u_n)_{n\in\mathbb{N}} \in A^{\mathbb{Z}}$ is linearly repetitive whenever there is $K > 1$ such that for each $n \in \mathbb{N}$ and all pair of finite subwords v, w with length $|v| = n, |w| = Kn$, then v occurs as a subword of w.

Theorem 4. *The tiling space of a linearly repetitive sequence is f-embeddable.*

It is possible to extend the notion of linear repetitivity to tilings in higher dimension [1]. The proof of the previous result, however, requires a stronger definition to be extended in higher dimension, related to forcing the border in a controlled way (See Section 6.3). This suggest that

Conjecture 1. *There is a suitable definition of linearly repetitive tiling in $\mathbb{R}^d$ such that the transversal of any such tiling space is f-embeddable*

In a similar way, the next consequence concerns Sturmian sequences. These are doubly infinite sequences such that number of words of length n it contains equals $n + 1$. A classical theorem (see, for instance, [5]) asserts that, for each such sequence $x = (x_n)_{n\in\mathbb{Z}}$, the alphabet contains two letters, usually denoted by $\{0, 1\}$ and there is an irrational number $\alpha \in (0, 1)$ such that $u_n = [x - n\alpha] - [x - (n + 1)\alpha]$ for some $x \in (0, 1]$. The following theorem is a corollary of the Main Result, and is proved in two parts: "if" (Proposition 12), and "only if" (Proposition 13).

Theorem 5. *The tiling space of a Sturmian sequence x associated with the irrational number $\alpha \in (0, 1)$ is f-embbedable if and only if α has bounded type, namely if and only if its continued fraction expansion has bounded partial quotients.*

A possible extension of this result is provided by the tiling spaces of *quasicrystals* obtained by the *cut-and-project* method [20]. Namely let $\mathbb{Z}^n$ seen as a subset of the Euclidean space $\mathbb{R}^n$. Let Δ be the unit cube $[0, 1]^n$ in $\mathbb{R}^n$. Let then $E_\parallel$ be a linear subspace of $\mathbb{R}^n$ of dimension d, with *irrational orientation*, namely such that $E_\parallel \cap \mathbb{Z}^d = \{0\}$. Then $\Sigma = \Sigma(E_\parallel) = (\Delta + E_\parallel) \cap \mathbb{Z}^d$. The set $\mathcal{L} = \mathcal{L}(E_\parallel)$ obtained as the orthogonal projection of $\Sigma(E_\parallel)$ in $E_\parallel$ gives a model for the atomic positions of a possible d-dimensional quasicrystal. The set $\mathcal{L}$ has several properties, in particular it is a Delone set of finite local complexity (FLC) [25]. A local patch is a finite subset of $E_\parallel$ of the form $(\mathcal{L} - x) \cap B(0; r)$ for some $x \in \mathcal{L}$. Here $B(0; r)$ denotes the closed ball centered at 0 with radius r. Then FLC means that, for each $r > 0$ the set of patches of radius r is finite. The *tiling space* of $\mathcal{L}$ is again the set of all Delone sets in $E_\parallel$, containing the origin sharing the same family of local patches. The combinatorial metric on the tiling space is defined in a similar way. Namely, if $\mathcal{L}$ and $\mathcal{L}'$ are two FLC Delone sets, then $d(\mathcal{L}, \mathcal{L}')$ is the inverse $1/R$ where R is the largest $r > 0$ such that the patch of radius r at the origin coincide in both $\mathcal{L}$ and $\mathcal{L}'$, namely such that $\mathcal{L} \cap B(0; r] = \mathcal{L}' \cap B(0; r]$. By analogy with the previous result the following conjecture is expected

Conjecture 2. *For Lebesgue almost all d-dimensional linear space $E_\parallel \in \mathbb{R}^d$, seen as points in the Grassmanian manifold of d-dimensional subspace of $\mathbb{R}^n$, the tiling space of $\mathcal{L}(E_\parallel)$, endowed with the combinatorial metric, is not f-embeddable.*

The last consequence analyzed here concerns random trees obtained from a Galton-Watson branching process [19, 3] (see [23] for a well documented fascinating history of this problem). Let p be a probability defined on the set $\mathbb{N} = \{0, 1, 2, \cdots\}$ of natural integers. The GW-tree is constructed inductively, starting from the root, as follows: if v is a vertex, it has M_v children where M_v is a random variable with distribution p and is independent from the M_u where u is any prior vertex constructed so far. It is a classical result in probability, partly proved initially by the Reverend Watson [35], called the *Galton-Watson-Haldane-Steffensen critical theorem* [23], that if the average number of children $\langle M_v \rangle$ is less than or equal to one, the tree obtained in this way is almost surely finite (extinction). In particular its reduced tree is empty. On the other hand, as was eventually proved by Steffensen in 1930, if $\langle M_v \rangle > 1$, the probability of extinction is less than one. However, the random tree produced in this way is likely to have dandling vertices.

One way to avoid such a property is to force every vertex to have at least two children. This can be done by demanding that $p_0 = p_1 = 0$, namely that p be supported by $[2, \infty) \subset \mathbb{N}$: then the Galton-Watson tree is automatically reduced and this model can be called a *Reduced Random Tree*. The first consequence of the Main Result is the following

Proposition 2. *Let p be a probability on the set $[2, \infty) \subset \mathbb{N}$. If it has an infinite support, then the number of children per vertex is almost surely unbounded. In particular, for any weight, the boundary of the reduced Galton-Watson tree associated with this probability is almost surely not f-embeddable.*

In such a case a *random weight* can be added in the following way: a family $(\lambda_v)_{v \in \mathcal{V}}$ of *i.i.d.* supported in $[0, 1]$, with common distribution ρ is defined. A weight is defined inductively by setting $\kappa(\emptyset) = 1$ and $\kappa(v) = \lambda_v \kappa(u)$ is v is a child of u. In order that this defines a weight, it is required that $\rho\{0\} = 0$ and $\rho\{1\} < 1$. Then

Theorem 6. *Let p be a probability on the set $[2, \infty) \subset \mathbb{N}$ and let ρ be a probability on $[0, 1]$ such that $\rho\{0\} = 0$ and $\rho\{1\} < 1$. If $m = \langle M_v \rangle < \rho\{1\}^{-1}$, let s_m be the unique solution of $\langle \lambda^s \rangle m = 1$. Then the Reduced Random Tree $\mathfrak{T}$ produced by the Galton-Watson process associated with p and endowed with the random weight associated with ρ gives rise to an ultrametric Cantor set $\partial \mathfrak{T}$ with Hausdorff dimension s_m almost surely. In addition, its Hausdorff measure exists almost surely and is a random probability measure.*

3. Necessary Conditions for f-Embeddability

The first result concerning the existence of an f-embedding is provided by

Proposition 3. *Let $\mathfrak{T}$ be a weighted Michon tree with weight κ such that $\partial \mathfrak{T}$ endowed with its ultrametric d_κ is f-embeddable. Then the number of children per vertex is uniformly bounded.*

Proof: By assumption, there is $n \in \mathbb{N}$, a constant $c \geq 1$ and a map $\phi : \partial \mathfrak{T} \to \mathbb{R}^n$ such that

$$(1) \qquad \frac{1}{c} d_\kappa(x, y) \leq \|\phi(x) - \phi(y)\| \leq c\, d_\kappa(x, y)\,, \qquad \forall x, y \in \partial \mathfrak{T}\,.$$

Let then $v \in \mathcal{V}$ be a vertex. By definition $\kappa(v)$ represents the diameter of $[v]$ in $\partial \mathfrak{T}$. Therefore $\mathrm{diam}\{\phi([v])\} \leq c\, \kappa(v)$. Moreover, if w, w' are two distinct children of v, the distance between $[w], [w']$ is equal to $\kappa(v)$ as well, thanks to the definition of d_κ. Therefore, given $x_w \in [w]$ for each child of v, the open balls $B(x_w, \kappa(v))$ are pairwise disjoint, because d_κ is an ultrametric. These balls are contained in $[v]$. Thanks to eq. (1) it follows that the euclidean open balls $B_w = B(\phi(x_w), \kappa(v)/c)$ are pairwise disjoint and contained in a Euclidean closed ball B_v of radius $c\kappa(v)$. Let M_v be the number of children of v. Let ω_n be the volume of the Euclidean ball of radius 1 in $\mathbb{R}^n$. Then the sum of the volumes of the B_w's is at most equal to the volume of B_v, leading to

$$M_v\, \omega_n \left(\frac{\kappa(v)}{c} \right)^n \leq \omega_n \left(c\, \kappa(v) \right)^n\,, \qquad \Rightarrow \qquad M_v \leq (c^2)^n\,,$$

$\square$

Remark 1. The Proposition 3 shows that if two Michon trees represents two equivalent ultra-metrics d, d' on the Cantor set C, then they both have a bounded number of children per vertex. $\square$

Proposition 4. *Let $\mathfrak{T}$ be a Michon tree with a weight κ and a uniformly bounded number of children per vertex. If the corresponding ultrametric Cantor set $(\partial \mathfrak{T}, d_\kappa)$ is f-embeddable, then there are $0 < \theta < 1$, and $c \geq 1$ such that for any pair (v, w) of vertices such that v is an ancestor of w*

$$\frac{\kappa(w)}{\kappa(v)} \leq c\, \theta^{|w| - |v|}\,.$$

Proof: Let M be the maximum number of children per vertex in $\mathfrak{T}$. It is finite by assumption and since $\mathfrak{T}$ is reduced, $M \geq 2$. Let $0 < \delta < 1$ be chosen. For each vertex $v \in \mathcal{V}$, let $\mathrm{Desc}_\delta(v)$ be the set of descendants w of v such that $\kappa(w) \leq \delta\, \kappa(v)$. This set is not empty since the weight converges to zero along any path passing though v. Let $S_\delta(v)$ be the maximal elements in $\mathrm{Desc}_\delta(v)$, so that $w \in S_\delta(v)$ if and only if w is a descendant of v such that $\kappa(w) \leq \delta\kappa(v)$ and its father $\hat{w}$ satisfies $\kappa(\hat{w}) > \delta\kappa(v)$.

Then a reduced graph $\mathcal{T}_\delta = (\mathcal{V}_\delta, \mathcal{E}_\delta, \emptyset, \kappa_\delta)$ is constructed as follows:

(i) its root is the same as the root of $\mathcal{T}$,

(ii) its set of vertices is a subset $\mathcal{V}_\delta \subset \mathcal{V}$ constructed inductively as follows:

    (a) $\mathcal{V}_{\delta,1} = S_\delta(\emptyset) \subset \mathcal{V}$,

    (b) $\mathcal{V}_{\delta,n+1} = \bigcup_{u \in \mathcal{V}_{\delta,n}} S_\delta(u)$,

(iii) an edge $e \in \mathcal{E}_\delta$ is a pair (v, w) where $v \in \mathcal{V}_\delta$ and $w \in S_\delta(v)$,

(iv) its weight is given by $\kappa_\delta(v) = \kappa(v)$ if $v \in \mathcal{V}_\delta$.

Let d_δ be the metric defined on $\partial \mathcal{T}_\delta$ by the weight κ_δ. It follows, from the construction of $\mathcal{T}_\delta$, that every infinite path $x \in \partial \mathcal{T}$ starting at the root defines also a unique infinite path $f(x) \in \partial \mathcal{T}_\delta$ starting at the same root, by selecting, among the vertices defining x, the ones belonging to $\mathcal{V}_\delta$. From this it follows that f is actually invertible and continuous. Hence, identifying $\partial \mathcal{T}$ and $\partial \mathcal{T}_\delta$ with the Cantor set C, f is just the identity map.

If x, y are distinct paths in $\partial \mathcal{T}$, then the vertex $x \wedge y \in \mathcal{V}$ may not belong to $\mathcal{V}_\delta$. However, if v is its least ancestor in $\mathcal{V}_\delta$, its weight satisfies always

$$\delta \kappa(v) < \kappa(x \wedge y) \le \kappa(v) \,.$$

Moreover, v is the least common ancestor of x and y inside the tree $\mathcal{T}_\delta$, so that $d_\delta(x, y) = \kappa(v)$ leading to

$$\delta \le \frac{d(x, y)}{d_\delta(x, y)} \le 1 \,.$$

Thus, the two metrics d, d_δ on the Cantor set C, are equivalent. In particular, (C, d) is embeddable if and only if (C, d_δ) is. Hence, thanks to Proposition 3, the maximum number M_δ of children per vertex in $\mathcal{T}_\delta$ is finite. It follows that the number of vertices in $S_\delta(v)$ is bounded by M_δ whenever $v \in \mathcal{V}_\delta$. Let then $r(v)$ be the maximal generation difference, in $\mathcal{T}$, between v and an element $w \in S_\delta(v)$. Since the path from v to w cannot contains more that M_δ points $r(v) \le M_\delta$.

Let v, w be two vertices in $\mathcal{V}$ such that $w \preceq v$. Let v_0 (resp. w_0) be the least ancestor of v (resp. w) belonging to $\mathcal{V}_\delta$. By construction w_0 is a descendant of v_0. If $v_0 = w_0$, then $\delta \kappa(v_0) < \kappa(w) \le \kappa(v) \le \kappa(v_0)$. Hence

$$\delta \; < \frac{\kappa(w)}{\kappa(v)} \le 1 \le \frac{\delta^{|w|-|v|}}{\delta^{M_\delta}} \,.$$

If now $w_0 \ne v_0$, then, by construction $\delta \kappa(w_0) < \kappa(w) \le \kappa(w_0)$, and $\kappa(w_0) \le \delta^n \kappa(v_0)$ where n is the generation difference between v_0 and w_0 in $\mathcal{T}_\delta$. Hence, the generation difference between v_0 and w_0, in $\mathcal{T}$, is at most nM_δ. Thus, if $|v_0|$ denotes the generation of v_0 in $\mathcal{T}$, $n \ge (|w_0| - |v_0|)/M_\delta$, so that

$$\frac{\kappa(w_0)}{\kappa(v_0)} \le \delta^{(|w_0|-|v_0|)/M_\delta} \,.$$

Since $|w| - |v| \le |w| - |v_0| \le |w| - |w_0| + |w_0| - |v_0| \le M_\delta + |w_0| - |v_0|$, this implies

$$\frac{\kappa(w)}{\kappa(v)} \le \frac{\kappa(w_0)}{\delta \kappa(v_0)} \le \frac{1}{\delta^2} \delta^{(|w|-|v|)/M_\delta} \,.$$

Since $\mathcal{T}$ is reduced, it follows that $M_\delta \ge M \ge 2$. Consequently, the conclusion holds with $c \le \delta^{-M_\delta}$ and if $\delta^{1/M_\delta} = \theta$. $\qquad\square$

4. The Embedding Theorem

While the previous section showed the the conditions (i) and (ii) in Theorem 3 are necessary, this Section will prove that they are sufficient. The proof is just and adjustment of the proof of Savinien and the second author in [22].

Proposition 5. *Let $\mathfrak{T}$ be a Michon tree with a weight κ with a bounded number of children per vertex and so that there are $c \geq 1$ and $0 < \theta < 1$ for which*

$$(2) \qquad \qquad \frac{\kappa(w)}{\kappa(v)} \leq c\,\theta^{|w|-|v|} \qquad \qquad \forall w \preceq v\,, \quad v, w, \in \mathcal{V}\,.$$

Then $(\partial\mathfrak{T}, d_\kappa)$ is f-embeddable.

Proof: (i) Let M be the maximum number of children per vertex in $\mathfrak{T}$. Let $g : \mathcal{V} \rightarrow \{1, 2, \cdots, M\}$ be a map such that its restriction to the set of children of v be one-to-one for all vertex $v \in \mathcal{V}$. Such a map consists simply in giving a numbering to each child of each vertex. Hence $1 \leq g(v) \leq M$ for all v. Let now L be a natural integer. The choice of L will be made later. Any point $x \in \partial\mathfrak{T}$ will be represented by the sequence $x = (v_n)_{n=0}^{\infty}$ of its vertices in $\mathfrak{T}$, with the convention that v_n has generation n. Then let $\phi : \partial\mathfrak{T} \rightarrow \mathbb{R}^L$ be defined by $\phi(x) = (\phi_r(x))_{r=1}^L$ with

$$\phi_r(x) = \sum_{j=0}^{\infty} g(v_{jL+r})\,\kappa(v_{jL+r-1})\,, \qquad \qquad 1 \leq r \leq L\,.$$

Since $g(v_n) \leq M$ and $\kappa(v_n) \leq c\,\theta^n$ it follows that the series converges absolutely and uniformly with respect to r and x.

(ii) Let now $x, y \in \partial\mathfrak{T}$ with $x \neq y$ with $x = (v_n)_{n=0}^{\infty}$ and $y = (w_n)_{n=0}^{\infty}$. Then let j_0, r_0 be natural integers such that $|x \wedge y| = j_0 L + r_0 - 1$ and $1 \leq r_0 \leq L$. It follows that the paths x and y share the same vertices until the generation $|x \wedge y|$ and they split at the next generation. Therefore,

$$\begin{aligned}
\phi_r(x) - \phi_r(y) \;=\; & \sum_{j=j_0+1}^{\infty} \left(g(v_{jL+r})\kappa(v_{jL+r-1}) - g(w_{jL+r})\kappa(w_{jL+r-1})\right) \\
& + \; \chi(r \geq r_0)\left(g(v_{j_0L+r})\kappa(v_{j_0L+r-1}) - g(w_{j_0L+r})\kappa(w_{j_0L+r-1})\right)
\end{aligned}$$

It should be remarked that

$$-g(w_{jL+r})\kappa(w_{jL+r-1}) \leq g(v_{jL+r})\kappa(v_{jL+r-1}) - g(w_{jL+r})\kappa(w_{jL+r-1}) \leq g(v_{jL+r})\kappa(v_{jL+r-1})\,,$$

leading to the inequality (see eq. (2)),

$$|g(v_{jL+r})\kappa(v_{jL+r-1}) - g(w_{jL+r})\kappa(w_{jL+r-1})| \leq M\,\kappa(x \wedge y)\,c\,\theta^{(j-j_0)L+r-r_0}$$

Therefore

$$(3) \qquad \qquad |\phi_r(x) - \phi_r(y)| \leq \frac{Mc}{1 - \theta^L}\,\kappa(x \wedge y)\,.$$

Since $\kappa(x \wedge y) = d_\kappa(x, y)$, this map is Lipshitz continuous. On the other hand, $\|\phi(x) - \phi(y)\| \geq |\phi_r(x) - \phi_r(y)|$ for all $1 \leq r \leq L$. In particular, using again eq. (3) and since $v_{j_0L+r_0} \neq w_{j_0L+r_0}$, it follows that

$$\|\phi(x) - \phi(y)\| \geq |\phi_{r_0}(x) - \phi_{r_0}(y)| \geq \kappa(x \wedge y) - \frac{Mc\,\theta^L}{1 - \theta^L}\,\kappa(x \wedge y)\,.$$

If L is chosen so that $\theta^L < (Mc + 1)^{-1}$, it follows that this map is actually bi-Lipshitz. $\square$

5. Hausdorff Dimension

The minimal value of the embedding dimension L found in the proof of Proposition 5 is not as good as what was found in [22] for self similar Cantor sets. This is because in this case, more restrictions are put on the weight, in particular a lower bound of a type similar to the upper bound given in eq. (2). In such a case it was shown that the minimal embedding dimension is the smallest integer larger than the Hausdorff dimension of the ultrametric Cantor set considered. This addresses the question of computing the Hausdorff dimension of the ultrametric Cantor set $(\partial\mathcal{T}, d_\kappa)$.

5.1. **Hausdorff Dimension: the main result.**

The definition of the Hausdorff dimension [16] starts with the following construction: given an open cover $\mathcal{U}$ of $C = \partial\mathcal{T}$ and given $s \in [0, \infty)$, let $\mathcal{H}^s(\mathcal{U})$ be defined by

$$\mathcal{H}^s(\mathcal{U}) = \sum_{U \in \mathcal{U}} \mathrm{diam}(U)^s\,.$$

In addition, let $\mathrm{diam}(\mathcal{U}) = \sup_{U \in \mathcal{U}} \mathrm{diam}(U)$. Then for $0 < \delta < 1$ let $\mathcal{H}^s_\delta(C)$ be defined by

$$(4) \qquad \mathcal{H}^s_\delta(C) = \inf_{\mathrm{diam}(\mathcal{U}) < \delta} \mathcal{H}^s(\mathcal{U})$$

It follows that $\delta' \leq \delta \Rightarrow \mathcal{H}^s_\delta(C) \leq \mathcal{H}^s_{\delta'}(C)$. In addition, if $\sigma > 0$ then $\mathcal{H}^{s+\sigma}_\delta(C) \leq \delta^\sigma \mathcal{H}^s_\delta(C) \leq \mathcal{H}^s_\delta(C)$. Consequently,
  (i) the limit $\lim_{\delta \to 0} \mathcal{H}^s_\delta(C) = \mathcal{H}^s(C)$ exists in $[0, +\infty) \cup \{+\infty\}$,
  (ii) there is a unique s_0 such that if $s > s_0$ then $\mathcal{H}^s(C) = 0$ whereas for $s < s_0$, $\mathcal{H}^s(C) = +\infty$. This unique value s_0 is precisely the Hausdorff dimension $s_0 = \dim_H(C, d)$. In order to give a better formula, the following definition will be nedded

Definition 2. *A finite subtree Γ of $\mathcal{T}$ is a tree graph $\Gamma = (\mathcal{V}_\Gamma, \mathcal{E}_\Gamma, \emptyset)$ where*
  (i) $\mathcal{V}_\Gamma \subset \mathcal{V}$ is finite and contains the root $\emptyset$,
  (ii) for every $v \in \mathcal{V}_\Gamma$ every vertex of the unique path $\gamma_v : \emptyset \to v$ in $\mathcal{T}$ belong to $\mathcal{V}_\Gamma$,
  (iii) $\mathcal{E}_\Gamma$ is the subset of edges in $\mathcal{E}$ of the form (v, w) where both v, w are in $\mathcal{V}_\Gamma$.
$\partial\Gamma$ will denote the set of endpoints. The finite subtree Γ is called full if for any vertex $v \in \Gamma \setminus \partial\Gamma$, each child of v is a vertex of Γ.

The main result of this Section is

Theorem 7. *Let $\mathcal{T}$ be a reduced Michon tree with weight κ. For any $\delta > 0$ let $\mathfrak{G}_\delta(\mathcal{T})$ be the set of all full finite subtrees of $\mathcal{T}$ such that $\max_{v \in \partial\Gamma} \kappa(v) < \delta$. Then*

$$\mathcal{H}^s_\delta(\partial\mathcal{T}) = \inf_{\Gamma \in \mathfrak{G}_\delta(\mathcal{T})} \sum_{v \in \partial\Gamma} \kappa(v)^s\,.$$
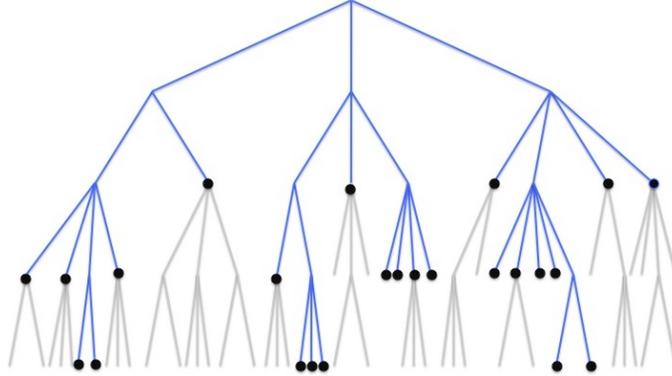
FIGURE 2. A full finite subtree and its extremal vertices

Corollary 1. *Let $\mathfrak{T}$ be a reduced Michon tree with weight κ. If the maximum number of children per vertex is M and if there are $c \geq 1$ and $0 < \theta < 1$ such that eq. (2) holds, then*

$$\dim_H(\partial\mathfrak{T}, d_\kappa) \leq \frac{\ln M}{|\ln \theta|} .$$

Given a vertex v let $K(v)$ be the product of $M(u)^{-1}$ over the ancestors of v. Then K defines a weight called the *Kraft weight*. This leads to the following elementary result

Corollary 2. *Let $\mathfrak{T}$ be a reduced Michon tree with its Kraft weight K. The Hausdorff dimension of $(\partial\mathfrak{T}, d_K)$ is exactly one. In addition, the Kraft weight defines a probability measure on $\partial\mathfrak{T}$ which coincides with the Hausdorff measure of $(\partial\mathfrak{T}, d_K)$. In particular $(\partial\mathfrak{T}, d_K)$ is f-embeddable if and only if $\mathfrak{T}$ has a bounded number of children per vertex.*

The proofs of the last three results will be the content of the following Sections.

5.2. **Partitions by Closed and Open Sets.**

Definition 3. *A c-partition of the Cantor set C is a finite family $\mathcal{P} = (P_1, P_2, \cdots, P_r)$ where the P_i's are closed and open (clopen) subsets of C with union C and with empty pairwise intersections.*

That c-partitions exist as a refinement of any open cover, is essentially equivalent to the definition of a Cantor set. This is because the topology admits a basis of clopen sets. Hence every cover admits a refinement made of clopen sets. By compactness, every clopen cover admits a finite subcover. At last, the intersection or the union of two clopen sets are clopen sets as well. Therefore the minimal elements of the σ-algebra generated by a finite clopen cover are themselves clopen and they make up a partition. This argument shows that any open cover admits a refining c-partition. Conversely a compact space with this property is completely disconnected and, if it has no isolated point, it is a Cantor set.

Proposition 6. *Let $\mathfrak{T}$ be a Michon tree with weight κ. The extremal vertices of a full finite subtree make up a c-partition of C. Conversely, for any c-partition $\mathcal{P}$ of the ultrametric Cantor set $(\partial\mathfrak{T}, d_\kappa)$, there is a full finite subtree Γ such that each element of $\mathcal{P}$ is a union of extremal vertices of Γ.*

Proof: (i) Let Γ be a full finite subtree of $\mathcal{T}$. If $v, w \in \partial\Gamma$ then neither of them is the descendant of the other, just because they are both extremal. Hence $[v] \cap [w] = \emptyset$. Since Γ is full, the brothers of any extremal vertex are all vertices in Γ. Let $x \in \partial\mathcal{T}$ be an infinite path $x = (v_n)_{n \in \mathbb{N}} \in \partial\mathcal{T}$. Since $v_0 = \emptyset$, there is a unique $n_0 \in \mathbb{N}$ such that $v_n \in \mathcal{V}_\Gamma$ for $n \leq n_0$ and $v_n \notin \mathcal{V}_\Gamma$ for $n > n_0$. If v_{n_0} were not an extremal vertex for Γ, then, since Γ is full, its child v_{n_0+1} should be in Γ as well, a contradiction. Therefore $v_{n_0} \in \partial\Gamma$ and $x \in [v_{n_0}]$. Consequently, the family $\{[v] \,;\, v \in \partial\Gamma\}$ is a c-partition.

(ii) The set $\{[v] \,;\, v \in \mathcal{V}\}$ is a basis for the topology of $\partial\mathcal{T}$. Given any clopen set $P \subset \partial\mathcal{T}$, any point $x \in P$ admits a neighborhood of the form $[v_x] \subset P$. The family $\{[v_x] \,;\, x \in \partial\mathcal{T}\}$ is an open cover of P. Since P is closed, it is compact so that a finite subcover can be extracted. But the intersection $[v] \cap [w]$ is either empty or, if not, one of the two $[v], [w]$ is included in the other. In the latter case, remove the smallest from the family. Proceeding this way, leads to a subcover which is actually a c-partition of P by vertices of the Michon graph. Let now $\mathcal{P}$ be a c-partition of $\partial\mathcal{T}$. Let $\mathcal{V}_\mathcal{P}$ be the set of vertices generating the c-partitions of the elements in $\mathcal{P}$. Let also Γ be the finite subgraph of $\mathcal{T}$ made with vertices and edges of the finite paths joining the root to one of the vertices in $\mathcal{V}_\mathcal{P}$. Then Γ is a finite subtree by construction. Actually, it is full: for indeed if not, it has an extremal vertex v having one of its brother w not in $\mathcal{V}_\Gamma$. But this in not possible because, $\mathcal{P}$ being a cover, $[w]$ is covered by $\mathcal{P}$ and thus has a non empty intersection with some $P \in \mathcal{P}$. Since $[w]$ does not intersect any of the $[u]$ with $u \in \mathcal{V}_\mathcal{P}$, there is a contradiction with the construction of $\mathcal{V}_\mathcal{P}$. Hence w must belong to $\mathcal{V}_\mathcal{P}$ as well. □

Lemma 1. *Let (X, d) be an ultrametric complete space. Then every open ball is closed and every closed ball is open.*

Proof: Let $x \in X$ and let $r > 0$. The ball $B(x; r) = \{y \in X \,;\, d(x, y) < r\}$ is open. If $z \in \overline{B(x;r)}$ then there is $y \in B(x; r)$ with $d(y, z) < r$. Therefore $d(x, z) \leq \max\{d(x, y), d(y, z)\} < r$ so that $z \in B(x; r)$. Conversely, let $\widehat{B}(x; r) = \{y \in X \,;\, d(x, y) \leq r\}$. If $y \in \widehat{B}(x; r)$ and if $z \in B(y; r)$ then

$$d(x, z) \leq \max\{d(x, y), d(y, z)\} \leq r \,,$$

so that $z \in \widehat{B}(x; r)$. In particular $B(y; r) \subset \widehat{B}(x; r)$ for all $y \in \widehat{B}(x; r)$. Hence $\widehat{B}(x; r)$ is open. □

Lemma 2. *Let (X, d) be an ultrametric compact space. Then for every open set $U \subset X$, there is a clopen set $P_U \supset U$ such that $\mathrm{diam}(P_U) = \mathrm{diam}(U)$.*

Proof: For $\delta > 0$ let $U^\delta = \{y \in X \,;\, \mathrm{dist}(y, U) < \delta\}$. By definition $U^\delta = \bigcup_{x \in U} B(x; \delta)$ which implies $\overline{U} = \bigcap_{\delta > 0} U^\delta$. In particular, U^δ is open. Since $\overline{U}$ is closed, it is compact and therefore there are $(x_1, \ldots, x_n)$ in U such that $\overline{U} \subset B(x_1; \delta) \cup \cdots \cup B(x_n; \delta) = P_U \subset U^\delta$. Since every open ball is also closed, it follows that P_U is clopen. Moreover $\mathrm{diam}(U) \leq \mathrm{diam}(P_U) \leq \mathrm{diam}(U^\delta)$. If now $x, y \in U^\delta$, there are $x', y' \in U$ such that $d(x, x') < \delta$ and $d(y, y') < \delta$. Therefore $d(x, y) \leq \max\{d(x, x'), d(x', y'), d(y', y)\} \leq \max\{\delta, \mathrm{diam}(U)\}$. In particular $\mathrm{diam}(U^\delta) \leq \max\{\delta, \mathrm{diam}(U)\}$. If $\delta < \mathrm{diam}(U)$, it follows that $\mathrm{diam}(P_U) = \mathrm{diam}(U)$. □

Lemma 3. *Let (C, d) be an ultrametric Cantor set. Then any finite cover by clopen set $\mathcal{P} = (P_1, P_2, \cdots, P_n)$ admits a refinement $\mathcal{Q} = (Q_1, Q_2, \cdots, Q_n)$ such that $Q_j \subset P_j$ and $\mathcal{Q}$ is a c-partition.*

Proof: The set $Q_1 = P_1 \setminus (P_2 \cup \cdots \cup P_n)$ is clopen and contained in P_1. It may happen that $Q_1 = \emptyset$. But in any case, $C = Q_1 \cup P_2 \cup \cdots \cup P_n$, while $Q_1 \cap P_j = \emptyset$ for $j > 1$. Hence replacing

C by $C_1 = C \setminus Q_1$ and $\mathcal{P}$ by $\mathcal{P}_1 = \mathcal{P} \setminus \{P_1\}$ the same argument follows. The Lemma is proved by induction. $\qquad \square$

5.3. **Proof of Theorem 7.** Let $\mathcal{V}$ be the set of vertices of the Michon tree $\mathcal{T}$. Thanks to the representation of c-partitions provided by the Proposition 6, each clopen set $P \subset \partial\mathcal{T}$ can be written as $P = \bigcup_{j=1}^{l} [v_j]$ where $(v_1, v_2, \cdots, v_l) \in \mathcal{V}$ are vertices which are pairwise not ancestor of each others. This means that $[v_i] \cap [v_j] = \emptyset$ if $v_i \neq v_j$. It follows that the least common ancestor $v_1 \wedge v_2 \wedge \cdots \wedge v_l = v_P$ provides a measure of the diameter of P namely

$$\mathrm{diam}(P) = \kappa(v_P), \qquad\qquad v_P = \bigwedge_{v \in \mathcal{V}\,;\,[v]\subset P} v\,.$$

Let now $\mathcal{P}$ be a c-partition $\mathcal{P}$. Several clopen sets of $\mathcal{P}$ might be associated with the same vertex v_P. In particular

$$\sum_{P \in \mathcal{P};\,v_P = v} \mathrm{diam}(P)^s \geq \kappa(v)^s\,.$$

In addition there might be clopen sets $P, P' \in \mathcal{P}$ such that v_P is an ancestor of $v_{P'}$. In such a case $\mathrm{diam}(P)^s + \mathrm{diam}(P')^s \geq \kappa(v_P)^s$. Let then $\mathcal{Q}' = \{[v_P]\,;\,P \in \mathcal{P}\}$. Since $[v_P] \supset P$, it follows that $\mathcal{Q}'$ is an open cover. In view of the previous remark, it might happen that $[v_P] \supset [v_{P'}]$. By removing the smaller set, it gives a subcover. Let then $\mathcal{Q}$ be the minimal subcover obtained from $\mathcal{Q}'$ in this way. Then $\mathcal{H}^s(\mathcal{P}) \geq \mathcal{H}^s(\mathcal{Q}') \geq \mathcal{H}^s(\mathcal{Q})$. In particular, $\mathcal{Q}$ is a c-partition made of *vertices* of the Michon graph. In addition, by construction, if $\mathrm{diam}(\mathcal{P}) < \delta$ it follows that $\kappa(v) < \delta$ for all $v \in \mathcal{V}$ such that $[v] \in \mathcal{Q}$. Hence the Proposition 6 implies that $\mathcal{Q}$ is defined by the extremal vertices of a full finite subtree $\Gamma \in \mathfrak{G}_\delta(\mathcal{T})$. $\qquad \square$

5.4. **Kraft's Identity.** Let $\mathcal{T} = (\mathcal{V}, \mathcal{E}, \emptyset, \kappa)$ be the reduced Michon tree of the ultrametric Cantor set (C, d). For each vertex $v \in \mathcal{V}$, let $M(v)$ be the number of its children. If $\gamma = (v_0 = \emptyset, v_1, \cdots, v_n)$ denotes a finite path starting at the root, then $M(\gamma)$ is defined by

$$M(\gamma) = \prod_{j=0}^{n-1} M(v_j)$$

Proposition 7 (see [9]). *If Γ is any full finite subtree of $\mathcal{T}$ then*

$$\sum_{\gamma \in \mathcal{P}_\Gamma} \frac{1}{M(\gamma)} = 1 \qquad\qquad\qquad \textbf{(Kraft's identity)}$$

where $\mathcal{P}_\Gamma$ denotes the set of all maximal paths in Γ starting at the root.

Proof: The proof is based upon an elementary remark: if $\gamma = (v_0, v_1, \cdots, v_{n-1}, v_n)$ is a path in $\mathcal{P}_\Gamma$, then the choice of v_n can only be made among the children of v_{n-1}. In particular, if $\mathrm{Ch}(v)$ denotes the set of children of v,

$$\frac{1}{M(v_{n-1})} \sum_{v_n \in \mathrm{Ch}(v_{n-1})} 1 = 1\,,$$

since $M(v_{n-1})$ is exactly the number of children of v_{n-1}. Since Γ is full, it follows that this identity applies also to paths in Γ. To vary γ further, it is necessary to sum over the possible

vertices v_{n-1}, namely among the children of v_{n-2}. Proceeding this way inductively gives the result. $\qquad\square$

Remark 2. The usual Kraft identity, in coding theory, deals with binary tree graphs, representing the possible strings of binary digits. Therefore $M(v) = 2$ for all vertex v. Then the previous identity reads $\sum_{\gamma\in\mathcal{P}_\Gamma} 2^{-|\gamma|} = 1$. Every maximal path can be identified with a code word. If no pair of code words share the same prefix, then the family of these code words lead to a finite subtree. This tree is full if the family of codewords is maximal among the set of codewords with no prefix in common. This prefix-free property makes this family of words recognizable when read. In particular a recognizable set of code words must be a subset of $\mathcal{P}_\Gamma$ for some full finite subtree of the binary tree. Hence if $\mathcal{W}$ is the set of such codewords and if $\ell(w)$ denotes the length of the word w, the following inequality is necessary for recognizability

$$\sum_{w\in\mathcal{W}} 2^{-\ell(w)} \le 1 \qquad\qquad \textbf{(Kraft's inequality)}$$

$\qquad\square$

5.5. **Proof of Corollary 1.** Let Γ be a full finite subtree of $\mathcal{T}$. Then, using the Kraft identity (see Proposition 7), it follows that

$$\sum_{v\in\partial\Gamma} \kappa(v)^s \le \sup_{v\in\partial\Gamma} \kappa(v)^s M(\gamma_v) \le c\,(\theta^s\,M)^{|v|}\,.$$

On the other hand, let N_δ be the smallest natural integer such that $c\theta^{N_\delta} \le \delta$. Then there is a full finite subtree Γ of $\mathcal{T}$ such that $v \in \partial\Gamma \;\Rightarrow\; |v| \ge N_\delta$. In particular, $\Gamma \in \mathfrak{G}_\delta(\mathcal{T})$ and therefore

$$\mathcal{H}_\delta^s(\partial\mathcal{T}) \le c\,(\theta^s M)^{N_\delta}\,.$$

Since $N_\delta \to \infty$ as $\delta \to 0$, it follows that if $\theta^s M < 1$, then $\mathcal{H}^s(\partial\mathcal{T}) = 0$, namely $s \ge \dim_H(\partial\mathcal{T})$. This gives the result.

6. S-ADIC SYSTEMS

6.1. **Definitions.** This Section is devoted to the basic definitions for S-adic systems. The presentation uses the notations of Durand [12, 13]. Let A be a finite alphabet. Note $A^\star$ the set of all finite words with letters in A. For $w \in A^\star$, $|w|$ denotes its length, namely the number of letters it contains. Consider S a *finite* set of morphisms

$$\sigma \in S : A(\sigma) \longrightarrow A^\star, \quad \text{with } A(\sigma) \subset A.$$

A S-adic system is a sequence $(\sigma_n : A_{n+1} \to (A_n)^\star)_{n\in\mathbb{N}} \in S^\mathbb{N}$, such that the morphisms are composable. It will be assumed that for all n, every letter in A_n appears in a word $\sigma_n(b)$ for some $b \in A_{n+1}$. For $m > n$ the following notation will be used

$$\sigma_{n,m} = \sigma_n \circ \ldots \circ \sigma_{m-1} : A_m \to A_n.$$

It satisfies $\sigma_{n,m} \circ \sigma_{m,k} = \sigma_{n,k}$. A S-adic system is *primitive* if there exists some $s_0 > 0$ such that for all $r \in \mathbb{N}$, for all $a \in A_{r+s_0}$ and all $b \in A_r$, the letter b appears in the word $\sigma_{r,r+s_0}(a)$. It is called *proper* if there exist two letters l and r in A such that for all $\sigma \in S$ and all $a \in A(\sigma)$, $\sigma(a)$ begins by the letter l and ends by r. Furthermore, it will be assumed from this point on, that

$$(5) \qquad\qquad \lim_{n \to +\infty} \min_{c \in A_n} |\sigma_{1,n}(c)| = +\infty,$$

Hence, words of arbitrary long length can be obtained from iterating the substitutions on a single letter. Given a S-adic system, there is a way to associate a subshift $\Xi \subset A^{\mathbb{Z}}$ (it is then called a S-adic subshift). Let T be the shift operator on $A^{\mathbb{Z}}$

$$T(\dots x_{-1} \cdot x_0 x_1 \dots) = \dots x_{-1} x_0 \cdot x_1 \dots.$$

It is straightforward to extend the morphisms of S to $A^{\mathbb{Z}}$ by concatenation

$$\sigma(\dots x_{-1} \cdot x_0 x_1 \dots) = \dots \sigma(x_{-1}) \cdot \sigma(x_0)\sigma(x_1) \dots.$$

It ought to be remarked that for all n, $\sigma_{1,n}(l)$ is a prefix of $\sigma_{1,n+1}(l)$. Similarly, $\sigma_{1,n}(r)$ is a suffix of $\sigma_{1,n+1}(r)$. Therefore, an element of $A^{\mathbb{Z}}$ can be defined by

$$x = \left(\lim_{n \to +\infty} \sigma_{1,n}(r) \right) \cdot \left(\lim_{n \to +\infty} \sigma_{1,n}(l) \right),$$

where the dot separates the x_i with $i \geq 0$ on the right from the ones with $i < 0$ on the left. The subshift associated with the S-adic system is the closure in $A^{\mathbb{Z}}$ (endowed with the product topology) of the orbit of x under the shift. This subshift, Ξ, is endowed with the combinatorial distance d, namely two sequences have distance $(n+1)^{-1}$ whenever they coincide on a string of radius n around the origin and do not coincide beyond. If the S-adic system is primitive, (Ξ, T) is minimal (see Durand [12, Lemma 7]).

Definition 4. *A word $x \in A^{\mathbb{Z}}$ is called linearly recurrent or linearly repetitive (LR for short) if there is a constant K such that for all $n \in \mathbb{N}$, for all subwords w and w' of x respectively of length n and Kn, then w occurs in w' as a subword.*

In other words[1], in a LR word, each finite subword repeat infinitely often and within a distance which varies linearly with their size. It is easy to see that a subshift generated by an LR word is minimal and that every elements of the subshift is LR with the same constant. In this case, the subshift itself is called *linearly recurrent*. The following is due to Durand

Theorem 8 (Durand [13], Proposition 1.1). *A subshift is S-adic primitive and proper if and only if it is linearly recurrent.*

The periodic case is trivial. In the non-periodic case, Durand's construction of the S-adic system associated with a linearly recurrent subshift involves return words. The S-adic system he builds in this case has the following property, which will be used in the proof of the result below (see [13, Section 4] for the construction, and [14, Definition 9, Lemma 17] for the properties of return words and codes).

Definition 5. *Let $(\sigma_n)_{n \in \mathbb{N}}$ be a S-adic system. It is said to have unique decomposition property if, given any element x in the subshift associated with Ξ, there is a unique decomposition of x as a concatenation*

$$x = \dots \sigma_1(a_{-1})\underline{\sigma_1(a_0)}\sigma_1(a_1)\dots,$$

where the index 0 of x is in the underlined word $\sigma_1(a_0)$.

---

[1]No pun intended.

6.2. **Proof of Theorem 4.** The periodic case is trivial (in this case, Ξ is finite). Using the theorem of Durand cited above, it is enough to prove the following

Proposition 8. *Let Ξ be a a primitive proper S-adic subshift with unique decomposition property and endowed with the combinatorial metric. Then, it is f-embeddable.*

It will be convenient to describe an S-adic subshift in terms of a Bratteli diagram (see for example [14]). A weight on the Bratteli diagram permits to define an ultrametric on the set of its infinite path. It will be necessary to prove

- the ultrametric Cantor set associated with this Bratteli diagram is bi-Lipschitz homeomorphic to the subshift with the combinatorial metric;
- the weight on the Bratteli diagram satisfies the hypothesis of Theorem 3 so that the Cantor set associated with the diagram is embeddable.

The following result is needed.

Lemma 4 (Durand [12], Lemma 8). *If the S-adic system generated by $(\sigma_n)_{n\in\mathbb{N}}$ is primitive with constant s_0, there exists a constant K such that for all integers r, s, with $s - r \geq s_0$ and for all b, c in A_{s+1}*

$$\frac{|\sigma_{r,s+1}(b)|}{|\sigma_{r,s+1}(c)|} \leq K.$$

Let $(\sigma_n : A_{n+1} \to A_n)$ be a primitive S-adic proper system with unique decomposition property and let l and r the letters associated with the properness. The Bratteli diagram is defined as follows

- For all $n \in \mathbb{N}$, the set of vertices $\mathcal{V}_n$ is in bijection with the alphabet A_n: for each $a \in A_n$, there is a $v_a \in V_n$.
- For all $n \in \mathbb{N}$, an edge in $\mathcal{E}_n$ is a triple $e = (v_a, l, v_b)$ where $v_a \in \mathcal{V}_n$, $v_b \in \mathcal{V}_{n+1}$ and $l \in \mathbb{N}$ such that the letter a occurs in the word $\sigma_n(b)$ in position $l + 1$. Hence, the number of edges from v_a to v_b is equal to the number of times the letter a appears in $\sigma_n(b)$.
- If $e = (v_a, l, v_b) \in \mathcal{E}_n$ its *source* is $s(e) = v_a$, its *range* is $r(e) = v_b$ and its *label* is $l(e) = l$.

Example 1. If $\sigma_n(b) = labcar$, then there are two edges e_1, e_2 from v_a to v_b: one corresponds to the second letter and the other to the fifth letter of the word above. Their respective labels are 1 and 4. □

Definition 6. *A path on the Bratteli diagram is a sequence of edges $(e_k)_{i\leq k<j}$ (with $1 \leq i < j \leq +\infty$), such that $e_k \in \mathcal{E}_k$ for all k and $r(e_k) = s(e_{k+1})$. Let Π_n denote the set of paths with $i = 1$, $j = n$ (simply called "paths of length n"). Let Π denote the union of all Π_n and let Π_∞ be the set of paths of infinite length ($i = 1$, $j = +\infty$).*

Because the σ_n's are taken from a finite set of substitutions S the cardinality of the sets $\mathcal{E}_n$ and $\mathcal{V}_n$ is uniformly bounded in n. Let

$$w_n = \big(\min\{|\sigma_{1,n}(a)| \ ; \ a \in A_n\}\big)^{-1}.$$

It is a decreasing sequence which tends to 0 as $n \to \infty$. The weight of a finite path γ will be defined by w_{n+1} whenever γ has length n. This leads to a metric on Π_∞ defined by

$$d_w(x, y) = w_{n+1}, \quad \text{where } n \text{ is the length of the longest common prefix of } x, y.$$

Proposition 9. *The subshift (Ξ, d), endowed with the combinatorial metric, and (Π_∞, d_w) are homeomorphic though a bi-Lipschitz homeomorphism.*

Proof: (i) **Constructing a map $\Xi \to \Pi_\infty$.** Let $x = \ldots x_{-1} \cdot x_0 x_1 \ldots$ in Ξ. Let $v_1 \in \mathcal{V}_1$ be the vertex corresponding to the letter x_0. By the unique decomposition property, x can be written in a unique way as the concatenation

$$x = \ldots \sigma(x'_{-1})\underline{\sigma(x'_0)}\sigma(x'_1) \ldots ,$$

with $(x'_i)_{i \in \mathbb{Z}} \in (A_2)^\mathbb{Z}$ and such that the letter of x of index 0 is in the underlined word $\sigma(x'_0)$. Therefore, if x' is the word $x' = (x'_i)_{i \in \mathbb{Z}}$

$$(6) \qquad\qquad\qquad\qquad x = T^k \sigma(x'),$$

This index k corresponds to an occurrence of the letter x_0 in the word $\sigma(x'_0)$. That is, it corresponds to an edge from v_1 to v_2, where $v_2 \in \mathcal{V}_2$ is the vertex corresponding to x'_0. Iterating this process of "de-substitution" leads to construct a sequence of words $x', x'', \ldots, x^{(n)} \ldots$ and a corresponding sequence of edges $e_1, e_2, \ldots, e_n$. In particular $\psi(x) = (e_1, e_2, \ldots)$ defines a map $\Xi \to \Pi_\infty$.

(ii) **ψ is a bijection.**   The map ψ has an inverse $\phi : \Pi_\infty \to \Xi$ which will be built explicitly. Let $\gamma = (e_1, e_2, \ldots)$ be a path going through the vertices $v_1, v_2, \ldots$. Given a finite word $w = (w_0, \ldots, w_{l-1}) \in A^*$, let ϵ be a symbol not in A (the "empty" symbol). Define then $\bar{w} \in (A \cup \{\epsilon\})^\mathbb{Z}$ is the sequence

$$\bar{w} = \ldots \epsilon \, \epsilon \, \epsilon \cdot w \, \epsilon \, \epsilon \, \epsilon \ldots .$$

This is a way of seeing a word in A^* as a (partially defined) element of $A^\mathbb{Z}$. For fixed n, the sequence

$$(7) \qquad \phi_n(\gamma) = T^{l(e_1)} \circ \sigma_1 \circ T^{l(e_2)} \circ \ldots \circ \sigma_{n-2} \circ T^{l(e_{n-1})+1} \left(\overline{r \, \sigma_{n-1}(a_n) \, l} \right) ,$$

where $l(e)$ denotes the label of the edge e, can be seen as an element of $(A \cup \{\epsilon\})^\mathbb{Z}$, namely as the sequence

$$\ldots \epsilon\epsilon \, \sigma_{1,n-1}(r) \, \underline{\sigma_{1,n}(a_n)} \, \sigma_{1,n-1}(l) \, \epsilon\epsilon \ldots ,$$

where the letter of index 0 occurs in the underlined word. Its exact position is determined by the labels of the edges. This implies

$$(8) \qquad\qquad [\phi_n(\gamma)]_i \neq \epsilon \quad \text{for} \quad -|\sigma_{1,n-1}(r)| \leq i \leq |\sigma_{1,n-1}(l)|.$$

By hypothesis, $\lim_{n \to +\infty} |\sigma_{1,n-1}(b)| = +\infty$ and this is true, in particular, whenever $b \in \{r, l\}$. Therefore, the limit as n tends to infinity of $\phi_n(\gamma)$ is an element of $A^\mathbb{Z}$, which is noted $\phi(x)$. Using the definitions of ϕ and ψ (see eq. (6) & (7)), it is straightforward that $\psi \circ \phi = \mathrm{id}_{\Pi_\infty}$. Therefore, ψ is onto and ϕ is one-to-one.

Conversely, let $x \in \Xi$, and $\gamma = \psi(x)$. It will be shown that $\phi(\gamma) = x$. For all n, x has a unique decomposition

$$x = \ldots \sigma_{1,n}(b_{-1})\underline{\sigma_{1,n}(b_0)}\sigma_{1,n}(b_1) \ldots ,$$

with $b_i \in A_n$. So by definition of ψ and ϕ,

$$(9) \qquad \phi_n(\gamma) = \ldots \epsilon \, \sigma_{1,n-1}(r) \, \underline{\sigma_{1,n}(b_0)} \, \sigma_{1,n-1}(l) \, \epsilon \ldots ,$$

and the two words coincide on the word $\sigma_{1,n}(b_0)$ (it appears at the same position). Furthermore, by properness of the S-adic system, $\sigma_n(b_{-1})$ ends with the letter r while $\sigma_n(b_1)$ begins with the letter l. Therefore, $\sigma_{1,n}(b_{-1})$ has $\sigma_{1,n-1}(r)$ as a suffix and $\sigma_{1,n}(b_1)$ has $\sigma_{1,n-1}(l)$ as a prefix. So for all $i \in \mathbb{Z}$ such that $[\phi_n(\gamma)]_i \neq \epsilon$, one has $[\phi_n(\gamma)]_i = x_i$. Taking a limit and using hypothesis (5), $\phi(\gamma)$ and x agree everywhere. Therefore, ϕ and ψ are inverse bijections of each other.

(iii) ϕ **is bi-Lipshitz.** Let $\gamma, \gamma' \in \Pi_\infty$ be such that the n first edges of γ and γ' coincide. Then, by definition, $\phi_n(\gamma) = \phi_n(\gamma')$. Thus $\phi(\gamma)$ and $\phi(\gamma')$ coincide for all indices i satisfying $-|\sigma_{1,n-1}(r)| \leq i \leq |\sigma_{1,n-1}(l)|$. Using Lemma 4 and the definition of w_n leads to

$$\forall b \in A_{n-1}, \qquad (w_{n-1})^{-1} \leq |\sigma_{1,n-1}(b)| \leq K(w_{n-1})^{-1}.$$

Since S is finite, $C = \max_{\sigma \in S, a \in A(\sigma)} |\sigma(a)|$ is well defined. Then, a word of the form $\sigma_{1,n}(b) = \sigma_{1,n-1} \circ \sigma_{n-1}(b)$ is at most C times longer than the longest word of the form $\sigma_{1,n-1}(c)$. That is, using again Lemma 4, $(w_n)^{-1} \leq CK(w_{n-1})^{-1}$. If $[\phi(\gamma)]_i = [\phi(\gamma')]_i$ for $|i| \leq K(w_{n-1})^{-1}$, then in particular this inequality holds for $|i| \leq (w_n)^{-1}/C$. So

$$d(\gamma, \gamma') \leq w_n \qquad \Rightarrow \qquad d(\phi(\gamma), \phi(\gamma')) \leq Cw_n.$$

Conversely, let γ and γ' coincide up to edge n, but differing on their $(n+1)$-th one. Then, by definition of ϕ

$$
\begin{aligned}
\phi(\gamma) &= \ldots \sigma_{1,n+1}(b_{-1})\sigma_{1,n+1}(b_0)\sigma_{1,n+1}(b_1)\ldots \\
\phi(\gamma') &= \ldots \sigma_{1,n+1}(b'_{-1})\sigma_{1,n+1}(b'_0)\sigma_{1,n+1}(b'_1)\ldots
\end{aligned}
$$

where $b_0 \neq b'_0$ and the letter of index 0 belongs to the word $\sigma_{1,n+1}(b_0)$ (respectively $\sigma_{1,n+1}(b'_0)$). By the unique decomposition property, $\phi(\gamma)$ and $\phi(\gamma')$ have to differ for indices i satisfying

$$-\max\{|\sigma_{1,n+1}(b_0)|, |\sigma_{1,n+1}(b'_0)|\} \leq i \leq \max\{|\sigma_{1,n+1}(b_0)|, |\sigma_{1,n+1}(b'_0)|\},$$

that is for

$$-K(w_n)^{-1} \leq i \leq K(w_n)^{-1}.$$

This proves that ϕ^{-1} is Lipschitz. $\qquad\square$

Proof of Proposition 9. In order to prove that (Ξ, d) is embeddable in $\mathbb{R}^p$, it is sufficient to prove that (Π_∞, d_w) is embeddable. It suffices to remark that the corresponding Michon tree has the finite paths in Π_∞ as vertices and edges given by pairs of path differing only by one edge. By the finiteness of S, all paths in Π_n have a bounded number of extensions to paths of Π_{n+1} and this bound is independent of n. Hence the Michon graph has a bounded number of children per vertex. So, the only thing left is to show that the sequence of weights $(w_n)_{n\in\mathbb{N}}$ on the Bratteli diagram is bounded above by a decreasing geometric sequence.

First, remark that for all n, the alphabet A_n has at least two elements. If one of the A_n had only one element, the subshift Ξ would be periodic (this case is already ruled out, because then the embedding is trivial). Using primitivity, for all letter $c \in A_{s_0+1}$, $|\sigma_{s_0}(c)| \geq 2$. By iteration, for all k and all $c \in A_{ks_0+1}$

$$|\sigma_{ks_0}(c)| \geq 2^k.$$

So $w_{ks_0} \leq 2^{-k}$. Since $(w_n)_{n \in \mathbb{N}}$ is decreasing, $w_n \leq C\lambda^n$ with $\lambda = 2^{-1/s_0} < 1$ and C is a constant. Thanks to Theorem 3 it follows that (Ξ, d) is f-embeddable.

6.3. **About the Conjecture 1.** The previous proof relies on the fact that linearly repetitive subshifts have a good representation by Brattli diagrams. While they are not self-similar, there are only a finite number of substitutions involved, which allows to generalize the methods of [22]. In a recent article, Aliste and Coronel [1] provide a good description of any linearly repetitive tiling space as the set of paths on a Bratteli diagram. It is possible in the linearly repetitive case, to have a Bratteli diagram such that the number of vertices and edges in each $\mathcal{V}_n$ or $\mathcal{E}_n$ is bounded uniformly in n. They also have estimates for the size of the patches associated with paths of size n. These estimates are analogues of Durand's Lemma 4 and allow to define reasonable weights on the Bratteli diagram. Therefore, the path space of this Bratteli diagram, endowed with the distance defined by the weight, is f-embeddable. However, there is no proof yet that the homeomorphism between this paths space and the tiling space is bi-Lipschitz.

In Proposition 9, the bi-Lipschitz character of ϕ is proved through using equation (9). This requires an estimate on the length of $\sigma_{1,n}(b_0)$, of $\sigma_{1,n-1}(r)$ and of $\sigma_{1,n-1}(l)$. This is a quantitative version of a property known as "forcing the border". It is unclear to the authors whether or not Aliste–Coronel's construction satisfies a similar quantitative border forcing property. Here is a possible analogue of the border forcing property that would be needed, restated in the context of [1].

Definition 7. *Let $(\sigma_n : A_{n+1} \to A_n)_{n \in \mathbb{N}}$ be a sequence of substitutions in $\mathbb{R}^d$ where A_i is a set of proto-tiles (compact subsets of $\mathbb{R}^d$ which are the closure of their interior). Let $C, \lambda > 1$ be two constant such that for all $n \in \mathbb{N}$ and all $t \in A_{n+1}$,*

$$C^{-1}\lambda^n \leq r_{\text{int}}(t) \leq R_{\text{ext}}(t) \leq C\lambda^n,$$

where $r_{\text{int}}(t)$ and $R_{\text{ext}}(t)$ are the inside and outside radius of the tile t, respectively. The system has quantitative border forcing property if there is a C' such that, for all n, all tile t, and all tiling T, if T contains $\sigma_{1,n}(t)$, then the patch of T contained in the $C'\lambda^n$-neighborhood of the patch $\sigma_{1,n}(t)$ does not depend on T.

This property means that the neighborhood of $\sigma_{1,n}(t)$ is forced, up to a distance which is controlled. In a recent article, Giordano-Matui-Putnam-Skau [18], can describe a tiling space as an inductive system (a sequence of "zoomed-out" tilings, in the sense of [4]), such that the size, and to some extend, the shape of the tiles involved is tightly controlled. In particular, the authors can avoid "flat tiles" in a certain sense. It is possible that their methods can be used to produce an inductive sequence associated with a tiling, which has a quantitative border forcing property. If this could be done in a way which is compatible with Aliste and Coronel's methods, the proof presented above could be adapted right away to the case of tilings of higher dimension. In the light of this discussion,it seems reasonable to expect that the Conjecture 1 could be proved using the ideas above. However, the technical difficulties, namely those induced by using GMPS's construction, would be quite important.

7. Sturmian Sequences

7.1. Definitions and notations. First, a few facts need to be introduced about Sturmian sequences and their coding. This section follows in part the presentation in [2], and proof for some of the results cited below can be found there.

Definition 8. *Given a sequence $x = (x_n)_{n \in \mathbb{Z}} \in A^{\mathbb{Z}}$ on the finite alphabet A, define its language:*

$$\mathcal{L}_n(x) = \{\textit{finite words } a_1 \ldots a_n \textit{ which appear in } x\}, \textit{ and } \mathcal{L}(x) = \bigcup_{n \in \mathbb{N}} \mathcal{L}_n(x).$$

Definition 9. *Given $x \in A^{\mathbb{Z}}$, its complexity function p_x is defined by*

$$\forall n \in \mathbb{N}, \quad p_x(n) = \mathrm{Card}(\mathcal{L}_n).$$

When there is no risk of ambiguity, $p_x(n)$ is just noted $p(n)$.
It is known that if for some n, $p_x(n) \leq n$, then the word x is periodic. Sturmian sequences can be defined using the complexity function: they are the aperiodic sequences which have the lowest possible complexity.

Definition 10. *A sequence $(x_n)_{n \in \mathbb{Z}}$ is called* Sturmian *if its complexity function satisfies $p(n) = n + 1$, and x is not eventually periodic.*

In particular, a Sturmian sequence is a sequence on two letters (since $p(1) = 2$). These two letters are noted 0 and 1.

Proposition 10. *The frequency of the letter 1 in a Sturmian sequence x*

$$\mathrm{freq}_x(1) := \lim_{n \to +\infty} \frac{\mathrm{Card}\{k \in [-n, n] \; ; \; x_k = 1\}}{2n + 1}$$

is well defined, and is an irrational number, noted $\alpha \in \,]0, 1[$.

Definition 11. *The Sturmian sequence x is of* type zero *if the frequency of 1 is less than one half, and of* type one *if this frequency is more than one half.*

It is easy to see that in a Sturmian sequence of type 1, the words 10, 01 and 11 may appear, but not the word 00. The same statements holds for sequences of type 0, just exchanging the letters 0 and 1.
Given a Sturmian sequence x, there is naturally a subshift of $\{0,1\}^{\mathbb{Z}}$ associated with it. It is by definition the set of all sequences y such that $\mathcal{L}(y) = \mathcal{L}(x)$. It can also be defined as the closure of the orbit of x in $A^{\mathbb{Z}}$ (for the product topology). The two definitions are equivalent. It is of course shift-invariant (hence the name "subshift"), and it is well known that it is minimal.
Note that if x has frequency of ones equal to α, so do all the elements of its subshift. In particular, it makes sense to write that a Sturmian subshift is of type 0 or of type 1. Conversely, the subshift associated with x is exactly the set of all Sturmian sequences which have the same frequency of ones. Therefore, it makes sense to denote $\Xi(\alpha)$ the subshift associated with $\alpha \in (0, 1)$.
Sturmian sequences can be *recoded* using the following substitutions. Define:

$$\sigma_0 : \begin{cases} 0 & \mapsto 0 \\ 1 & \mapsto 10 \end{cases} \quad \text{and} \quad \sigma_1 : \begin{cases} 0 & \mapsto 01 \\ 1 & \mapsto 1. \end{cases}$$

Proposition 11. *For any Sturmian sequence x of type 0, there is a Sturmian sequence x' such that either $x = \sigma_0(x')$ or $x = T\sigma_0(x')$. (T is the shift operator, see Section 6.)*
For any Sturmian sequence x of type 1, there is a Sturmian sequence x' such that either $x = \sigma_1(x')$ or $x = T\sigma_1(x')$.

Note Φ this re-coding map $x \mapsto x'$. It is worth noting that if x, y are two elements of the same Sturmian subshift, then $\Phi(x)$ and $\Phi(y)$ belong to the same subshift. Non-periodicity of Sturmian sequences implies that for any Sturmian sequence y of type 0 (resp. 1), there is a k such that $\Phi^k(y)$ is of type 1 (resp. 0).

Definition 12. *Let x be a Sturmian sequence, and $(\Phi^n(x))_{n \in \mathbb{N}}$ be the sequence of re-coded Sturmian sequences. By definition, for all n,*

$$\sigma_0^{b_1} \circ \sigma_1^{b_2} \circ \ldots \circ \sigma_0^{b_{2n+1}}(\Phi^{2n+1}(x)) \text{ and } \sigma_0^{b_1} \circ \sigma_1^{b_2} \circ \ldots \circ \sigma_1^{b_{2n}}(\Phi^{2n}(x))$$

are in the same orbit as x. All the b_n are positive, except maybe $b_0 = 0$. The sequence $(b_n)_{n \in \mathbb{N}}$ is called the multiplicative coding of the Sturmian sequence x.

All Sturmian sequences in a same subshift have the same multiplicative coding, and an acceptable multiplicative coding determines uniquely a Sturmian subshift.

7.2. **Partial fraction decomposition and multiplicative coding.** The properties of a Sturmian subshift are closely related to the partial fraction decomposition of the number α (which is the frequency of ones in the subshift).

Let $\alpha \in \mathbb{R}$ be irrational. Then α can be written uniquely $\alpha = a_0 + \alpha_0$, with $a_0 \in \mathbb{Z}$ and $\alpha_0 \in (0, 1)$. The Gauss map $G : [0, 1] \to [0, 1]$, applied to α_0, generates the continuous fraction expansion

$$G(\alpha) = \frac{1}{\alpha} - a(\alpha), \qquad\qquad a(\alpha) = \left[\frac{1}{\alpha}\right],$$

where $[x]$ denotes the *integer part* of x, namely the largest integer smaller than or equal to x. Hence

$$\alpha = a_0 + \frac{1}{a_1 + \alpha_1}, \qquad\qquad a_1 = a(\alpha), \ \ \alpha_1 = G(\alpha).$$

Iterating this formula gives rise to the continuous fraction expansion

$$\alpha = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots + \cfrac{1}{a_n + \alpha_n}}}}, \qquad\qquad \alpha_{n+1} = G(\alpha_n), \ a_n = a(\alpha_{n-1}) \ \ n \geq 1.$$

The standard notation is

$$\alpha = [a_0; a_1, a_2, \cdots, a_n, \cdots],$$

and the a_n's are called the *partial quotients* of α.

Definition 13. *A number α has* bounded type *whenever, the sequence of its partial quotients is bounded.*

Having bounded type is an exceptional property. This is a theorem of Khintchine. See also Levy [26].

Theorem 9 (see [24]). *For almost every $\alpha \in [0, 1]$ the sequence $(a_n)_{n \in \mathbb{N}_*}$ of partial quotients of α is unbounded.*

One famous example of such a typical number is $e = 2,71828\cdots$ the continued fraction of which was computed by Leonhard Euler in 1737 [15] namely

$$e = [2, 1, 2, 1, 1, 4, 1, 1, 6, \cdots, 1, 1, 2l, \cdots] \qquad\qquad a_{3l-1} = 2l, \ \ a_{3l-2} = a_{3l} = 1, \ \ l \geq 1.$$

Many properties of a Sturmian subshift are determined by the arithmetic properties of the number α associated with it. The following theorem is proved in Hedlund and Morse's seminal paper.

Theorem 10 (see [29]). *Consider a Sturmian subshift $\Xi \subset \{0,1\}^{\mathbb{Z}}$, of parameter α. This subshift is* linearly repetitive *(see definition 4) if and only if α has bounded type.*

An immediate consequence of this theorem and theorem 4 is the following result.

Proposition 12. *Let Ξ be a Sturmian subshift with parameter α, endowed with the combinatorial metric. If α has bounded type, then Ξ is f-embeddable.*

This result needs a converse statement: if α has not bounded type, then the associated Sturmian subshift is not f-embeddable. This will be proved in section 7.3.
One way to understand the deep links between the combinatoric properties of a Sturmian subshift and the arithmetic properties of its parameter α is the following. A Sturmian sequence x can be seen as the coding of an orbit of the rotation of angle α on the circle: there is $\{I_0, I_1\}$ a partition of the circle, and s a point in the circle such that $x_n = 0$ if and only if $(s + \alpha \mod 1) \in I_0$. The rotation on the circle can be related to the partial fraction decomposition of α on the one hand, and on the multiplicative coding of x on the other hand, to get to following result.

Theorem 11 ((see [2]). *Let x be a Sturmian sequence, and $\alpha = [0; a_1, a_2, \ldots]$ be the frequency of 1 in x. Then the multiplicative coding of x is given by the partial quotients of $(1-\alpha)/\alpha = \alpha^{-1} - 1$.*

It is straightforward that $(1 - \alpha)/\alpha = [a_1 - 1; a_2, a_3, \ldots]$. In particular, α has bounded type if and only if the sequence of coefficients of the multiplicative coding is bounded.

7.3. **Non embeddability of certain Sturmian subshifts.** This section is devoted to the proof of the following result.

Proposition 13. *Consider a Sturmian sequence x, with associated subshift Ξ and associated parameter α. If α has unbounded partial quotients, then Ξ (with the combinatorial metric) is not embeddable in a finite dimensional space.*

A Sturmian subshift Ξ is well described combinatorially by the (bilateral) tree of words of its elements.

Definition 14. *Given a Sturmian subshift Ξ, its un-reduced tree of words is defined as follows.*
 – *For all $n \geq 0$, define a sequence of refining partitions of Ξ by:*
$$\mathcal{P}_n = \{[y_{-n} \ldots y_n] \ ; \ y \in \Xi\},$$
 where $[y_{-n} \ldots y_n]$ is the cylinder set of all words in Ξ which coincide with y on indices $-n \leq i \leq n$.
 – *The set of vertices of the tree is in bijection with the disjoint union of all the $\mathcal{P}_n$. If $X \in \mathcal{P}_n$, the associated vertex is noted v_X.*
 – *There is an edge between v_X and v_Y if and only if $X \in \mathcal{P}_n$, $Y \in \mathcal{P}_{n+1}$ and $Y \subset X$.*
 – *Add a vertex v_Ξ, and an edge between v_Ξ and the vertices v_X for $X \in \mathcal{P}_0$.*
 – *The weight of the vertex v_X is $1/(n+1)$ if $X \in \mathcal{P}_n$ $(n \geq 0)$, and 1 otherwise.*

The (reduced) tree of words is obtained from this tree by the reduction process defined in section 2.

One remark about this tree: if $X \in \mathcal{P}_n$, then $\mathrm{diam}(X) \leq n$, for the combinatorial distance, with equality if and only if X is the non-trivial union of two distinct elements of $\mathcal{P}_{n+1}$. Note that the vertices v_X of such clopen sets X are exactly the vertices with two children: these are precisely the ones which are not dropped by the reduction process.

This leads to the following proposition.

Proposition 14. *The boundary of the reduced tree of words associated with a Sturmian subshift Ξ is bi-Lipschitz homeomorphic to the subshift (endowed with the combinatorial metric).*

Proof: The proof is almost tautological: given an infinite path in the tree, say γ, the sequence of vertices $(v_{X_n})_{n \in \mathbb{N}}$ defines a decreasing sequence of compact sets X_n, the diameter of which tends to zero. Therefore, its intersection is not empty and consists of a single element $\{x\}$. Define $\phi(\gamma) = x$.

Conversely, given $x \in \Xi$, there is a unique decreasing sequence of sets $X_n \in \mathcal{P}_n$, such that for all n, $x \in X_n$ (explicitly: $X_n = [x_{-n} \ldots x_n]$). Then it defines an infinite path $\gamma \in \partial T$. Clearly, γ is the unique pre-image of x by ϕ.

The fact that ϕ is bi-Lipschitz (and in particular, it is a homeomorphism) results from the remark above, on the diameters of the elements of $\mathcal{P}_n$. $\qquad\qquad\square$

Lemma 5. *Let y be a Sturmian sequence, assume that $y = \sigma_0^{b_n}(z)$, with z a Sturmian sequence. Then y contains the words $10^{b_n}1$ and $10^{b_n}0$.*

Proof: If $y = \sigma_0^{b_n}(z)$, then y is of type 0 and z is of type 1. By iteration, it is straightforward that $\sigma_0^{b_n}(0) = 0$ and $\sigma_0^{b_n}(1) = 1(0^{b_n})$. Since z is of type 1, it contains the words 10 and 11. Therefore, y contains the words:

$$1(0^{b_n})1(0^{b_n}) \quad \text{and} \quad 1(0^{b_n})0.$$

$\qquad\qquad\square$

Proof of Proposition 13: Let x be a Sturmian sequence, with α the frequency of 1, and assume that α does not have bounded type. Then its multiplicative coding $(b_n)_{n \in \mathbb{N}}$ is an unbounded sequence.

Let $y = \Phi^{b_0 + b_1 + \ldots + b_{n-1}}(x)$, and $z = \Phi^{b_n}(y)$. Without loss of generality, we assume that z is of type 1, so that

$$y = \sigma_0^{b_n}(z).$$

Then, using previous lemma, y contains the words $10^{b_n}1$ and $10^{b_n}0$. Therefore, for all $1 \leq k \leq [b_n]/2 - 1$, y contains the words $00^{2k}0$ and $00^{2k}1$.

Applying the substitution $\sigma := \sigma_0^{b_1} \circ \ldots \circ \sigma_0^{b_{n-1}}$, the sequence x contains the words

$$\sigma(0)\sigma(0)^{2k}\sigma(0) \quad \text{and} \quad \sigma(0)\sigma(0)^{2k}\sigma(1).$$

Let a be the last letter of $\sigma(0)$ and $\sigma(1)$ (it is the same: 0 if $b_0 \neq 0$, 1 otherwise). Since $\sigma(1)$ starts by 1 and $\sigma(0)$ starts by 0, x contains the words

$$(10) \qquad\qquad a\sigma(0)^k\sigma(0)^k0 \quad \text{and} \quad a\sigma(0)^k\sigma(0)^k1.$$

Then, let $X_k = [a\sigma(0)^k \cdot \sigma(0)^k]$, where the dot separates the indices $i \leq 0$ and $i > 0$. It is an element of $\mathcal{P}_{k|\sigma(0)|}$, where $|\sigma(0)|$ is the length of $\sigma(0)$. Let v_k be the associated vertex. From Equation (10), the vertices v_k have two distinct children (in the un-reduced tree of words), therefore, they are elements of the reduced tree, and their weight is $(k|\sigma(0)|)^{-1}$. In particular,

this construction shows that there are two vertices u, v, such that the quotient of their weights is $[b_n/2] - 1$, and their distance in the tree is $[b_n/2] - 1$.

This construction can be done for all n. If $(b_n)_{n \in \mathbb{N}}$ is unbounded, this shows that the weights cannot satisfy the geometric decay condition of Theorem 3, and Ξ is not embeddable.   □

The results presented here (namely proposition 12 and 13) provide a proof of Theorem 5. It is actually possible to give a more precise version of it.

Theorem 12. *A Sturmian subshift $\Xi(\alpha)$ is f-embeddable if and only if the irrational number α associated with it has bounded type. In particular*

   (i) if α is a quadratic irrational, then $\Xi(\alpha)$ is f-embeddable;
   (ii) for almost every $\alpha \in (0, 1)$, the subshift $\Xi(\alpha)$ is not f-embeddable;
   (iii) the boundary of $\Xi(e)$ is not embeddable for $e = 2.71828\ldots$.

Proof: The first part of this result is theorem 5. Point (i) is a consequence of the fact that quadratic irrational have an eventually periodic (hence bounded) partial fraction expansion. Point (ii) is a consequence of the theorem of Khintchine on continued fractions, and point (iii) is a consequence of the explicit formula for the partial fraction decomposition of e.   □

8. Random Trees

Random trees are good candidates to provide examples of Cantor sets with typical properties. In the history of random processes, the Galton-Watson process has certainly be a seminal example. Initially, Sir Francis Galton was concerned by the possible disappearance of family names among the British aristocrats in the nineteen century. He addressed this problem in a published article. Soon after, the reverend Watson proposed a method to attack the problem and provided a solution [35]. Unfortunately his prediction was that family names where all disappearing with probability one, which was certainly not what Galton could observe. Galton encouraged Watson to search further. It took up to the 1930's and the work of Haldane and Steffensen before the complete solution was found (see [23] for a fascinating review). As it turns out, branching processes of the Galton-Watson type appears in many other situations. For instance the description of the nuclear chain reaction or of electron emissions in a photomultiplier tubes can be described faithfully by such a process (see [19]).

The Galton-Watson process, as it is called today [19, 3] starts from a root and defines the number $\xi_\emptyset$ of offspring as an integer valued random variable with distribution p. Each offspring will be represented as a vertex v of the first generation of the random tree and a new integer valued random variable ξ_v, describing the number of offspings of v, is drawn with probability p so that all the ξ_v's, including the root, are stochastically independent. Each vertex v such that $\xi_v \neq 0$, will have ξ_v offsprings represented as vertices of the second generation. Proceeding in the same way for all descendants of the root, the branching process is built inductively. The main difficulty, for the purpose of the present paper, is that the resulting random tree has, in general, a non zero probability to be finite. Moreover, even if it is infinite, it has an infinite number of dead branches, namely finite maximal paths. This is because, any vertex has a nonzero probability to have no offsprings or to have only one. The problem of pruning the tree to reduce it, leads to some difficulties that are interesting from the probabilistic point of view but provides useless complications as a nice illustration of the core result of the present article. For this reason, the model will be simplified by requiring that the probabilities p_0 and p_1, that the number of offspring be zero or one, respectively, vanish. So, from now on *with probability one, the number*

of offsprings will be at least two. In this way, each vertex has at least two children, so that the resulting tree is already reduced. This is why this branching process will be called a *Reduced Random Tree.*

8.1. **Reduced Random Trees: a review.** Let $p = (p_n)_{n \geq 2}$ be the probability supported by $[2, \infty) \subset \mathbb{N}$ describing the number of children of each vertex. Let Z_n be the number of descendant at the generation n. Following Watson's idea, it is convenient to introduce the generating function

$$P_n(x) = \sum_{l=2}^{\infty} \text{Prob}\{Z_n = l\} \, x^l$$

The construction of the random tree implies the following formula, where $\mathcal{V}_n$ denotes the set of vertices at generation n,

$$(11) \qquad\qquad Z_{n+1} = \sum_{v \in \mathcal{V}_n} \xi_v \, .$$

Consequently the conditional probabilities are given by

$$\text{Prob}\{Z_{n+1} = l \,|\, Z_n = k\} = \sum_{j_1 + \cdots + j_k = l} p_{j_1} \cdots p_{j_k} \, .$$

In particular, it shows that $(Z_n)_{n \in \mathbb{N}_*}$ defines a Markov chain. Moreover

$$(12) \qquad\qquad P_{n+1}(x) = P_n\big(P(x)\big) , \qquad\qquad P(x) = \sum_{n=2}^{\infty} p_n \, x^n \, .$$

Since p is a probability, the series defining P converges for $0 \leq x \leq 1$. In addition, $P(1) = 1$, $m = P'(1) = \mathbb{E}(\xi)$ represents the average number of offsprings, if it exists. With the present restrictions, it follows that $m \geq 2$. The function $x \in [0, 1] \to P(x) \in [0, 1]$ is positive, increasing and convex, more generally all its derivatives, when they exist, are positive, meaning that P is *completely monotone.* From the recursion relation (12) it follows immediately that

$$P_n(x) = \underbrace{P \circ P \circ \cdots \circ P}_{n}(x)$$

In particular, since $P(1) = 1$, it follows that $\mathbb{E}(Z_n) = P_n'(1) = m^n$ for all n. Hence the number of descendants at generation n grows exponentially fast with the generation in the average.

Let now $\mathcal{F}_n$ be the sigma algebra generated by the variables ξ_v for v vertices of the generations $k \leq n$. A classical remark made by Doob [11], is that, thanks to the equation (11) defining the process, the family $(Z_n)_{n \in \mathbb{N}_*}$ satisfies

$$\mathbb{E}\left(Z_{n+1} | \mathcal{F}_n\right) = m Z_n \, .$$

Namely $W_n = Z_n / m^n$ is a *martingale.* As shown earlier by Doob [10] this implies

Theorem 13 (See [19, 3]). *Let $p = (p_n)_{n \geq 2}$ be the probability distribution for the number ξ of offsprings such that the average $\mathbb{E}(\xi) = m$ and the variance $\text{Var}(\xi) = \sigma^2$ are finite. Then the sequence $W_n = Z_n / m^n$ of random variables is a martingale with respect to the increasing*

sequence $\mathcal{F}_n$ of σ-algebras. In particular it converges almost surely to a random variable W such that

$$\mathbb{E}(W) = 1, \qquad\qquad \mathrm{Var}(W) = \frac{\sigma^2}{m^2 - m}.$$

8.2. **Proof of the Proposition 2.** If the probability p has an infinite support, given any integer $M \geq 2$, the probability P_M that a given vertex has more than M children is non zero. The construction of the Random Reduced Tree, can be seen by associating inductively with any vertex of generation n a string $(b_1, b_2, \cdots, b_n)$ of integers so that $1 \leq b_n \leq \xi_{b_1, \cdots, b_{n-1}}$. In particular, since $\xi_v \geq 2$ for all v's almost surely the subset $\mathcal{W} \subset \mathcal{V}$ made of vertices for which all b_j's belong to $\{1, 2\}$ is non empty and gives an infinite binary subtree. Since the random variables $\{\xi_v ; v \in \mathcal{W}\}$ are *i.i.d.*, it follows that, given $M \geq 2$

$$\mathrm{Prob}\{\xi_v \leq M ; \forall v \in A\} = (1 - P_M)^{\#A}.$$

In particular the probability that all vertices in $\mathcal{W}$ have $\xi_v \leq M$ vanishes. $\qquad\square$

8.3. **Random Weight.** In order that the boundary of the rooted random tree $\mathcal{T} = (\mathcal{V}, \mathcal{E}, \emptyset)$ previously built becomes an ultrametric Cantor set, it is necessary to put a weight on each vertex. The previous construction suggests that the weight itself be random and Markovian as well. In order to do so, the following model of random weight is proposed: let $(\lambda_v)_{v \in \mathcal{V}}$ be a family of *i.i.d* in $[0, 1]$ with common distribution ρ; then the weight $\kappa(v)$ will be given by

$$\kappa(v) = \lambda_v \kappa(u), \qquad \text{if } u = \text{father of } v, \qquad\qquad \kappa(\emptyset) = 1.$$

In order that this defines a good weight, it is required that $\lambda_v \neq 0$ with probability one, namely $\rho\{0\} = 0$. Moreover, in order that the weight converges to zero along any infinite paths with probability one, it will be required that $\rho\{1\} < 1$. It will be convenient to use the following generating function (Mellin transform)

$$h(s) = \int_0^1 \lambda^s \rho(d\lambda).$$

It is easy to see that h is decreasing, logarithmically convex, namely $h(\eta s_1 + (1 - \eta)s_0) \leq h(s_1)^\eta h(s_0)^{1-\eta}$ for $0 < \eta < 1$, and $\lim_{s \to \infty} h(s) = \rho\{1\}$.

8.4. **Proof of Theorem 6.** Thanks to Theorem 7, the Hausdorff dimension of the tree can be computed from the following random variables

$$(13) \qquad \mathcal{H}^s(\Gamma) = \sum_{v \in \partial\Gamma} \kappa(v)^s, \qquad\qquad \mathcal{H}_n^s = \sum_{v \in \mathcal{V}_n} \kappa(v)^s.$$

where Γ is a full finite subtree of the random $\mathcal{T}$. In the following $\mathfrak{G}$ will denote the set of finite full subtrees of $\mathcal{T}$. This set is ordered by the inclusion of the vertex sets. In addition, giving $\Gamma, \Gamma' \in \mathfrak{G}$, then they admit a least upper bound $\Gamma \vee \Gamma'$ and a greatest lower bound $\Gamma \wedge \Gamma'$, as can be checked immediately. This suggests to define a new family of σ-algebras: let $\widehat{\mathcal{F}}_\Gamma$ be the σ-algebra generated by the ξ_u's and the λ_v's where $v \in \mathcal{V}_\Gamma$ and $u \in \mathcal{V}_\Gamma \setminus \partial\Gamma$. In particular, if Γ is the full tree associated with generation n, then $\widehat{\mathcal{F}}_n$ will denote the σ-algebra generated by the ξ_u's with u vertex of generation $k \leq n - 1$ and by the λ_v's with v vertex of generation $k \leq n$. With these notations, the following result holds

Proposition 15. *With the assumptions made previously on the construction of the random rooted weighted tree $(\mathfrak{T}, \mathcal{V}, \emptyset, \kappa)$ the following results hold*
(i) For all $s \geq 0$, the family $Y_n(s) = \mathcal{H}_n^s/m^n h(s)^n$ is a martingale with respect to the family $\left(\widehat{\mathcal{F}}_n\right)_{n \in \mathbb{N}}$ of σ-algebras. In particular it converges almost surely to a positive random variable $Y(s)$ such that $\mathbb{E}(Y(s)) = 1$.
(ii) There is $t_m > s_m$ defined as the unique solution of $h(2s) = mh(s)^2$, such that
   (a) if $s < t_m$ then $Y_n(s)$ converges almost surely to a constant,
   (b) if $s = t_m$, then,

$$\mathrm{Var}(Y(t_m)) = 1 - \frac{1}{m} + \frac{\sigma^2}{m^2},$$

(c) if $s > t_m$ the random variable $Y(s)$ has not a finite second moment.
(iii) If $\rho\{1\} < m^{-1}$ and if s_m is the unique solution of $mh(s) = 1$, then the sequence of random variables $Y(\Gamma) = \mathcal{H}^{s_m}(\Gamma)$ defines a martingale with respect to the family $\left(\widehat{\mathcal{F}}_\Gamma\right)_{\Gamma \in \mathfrak{G}}$ of σ-algebras. In particular it converges almost surely to 1.

Proof: (i) From the definition of $\mathcal{H}_n^s$ in eq. (13), it follows that

$$\mathbb{E}\left(\mathcal{H}_1^s\right) = \mathbb{E}\left(\sum_{v \in \mathrm{Ch}(\emptyset)} \lambda^s\right) = h(s)\mathbb{E}(\xi_\emptyset) = mh(s).$$

Moreover,

$$
\begin{aligned}
\mathbb{E}(\mathcal{H}_{n+1}^s \,|\, \widehat{\mathcal{F}}_n) &= \mathbb{E}\left(\sum_{u \in \mathcal{V}_n} \kappa(u)^s \sum_{v \in \mathrm{Ch}(u)} \lambda_v^s \,\Big|\, \widehat{\mathcal{F}}_n\right) \\
&= h(s)\mathbb{E}\left(\sum_{u \in \mathcal{V}_n} \kappa(u)^s \xi_u \,\Big|\, \widehat{\mathcal{F}}_n\right) = m\, h(s) \sum_{u \in \mathcal{V}_n} \kappa(u)^s \\
&= mh(s)\mathcal{H}_n^s.
\end{aligned}
$$

In particular, this calculation shows that $\mathbb{E}(Y_{n+1}(s) \,|\, \widehat{\mathcal{F}}_n) = Y_n(s)$. In particular it is a martingale w.r.t. the family $\widehat{\mathcal{F}}_n$ of σ-algebras. Since $Y_1(s) = \mathcal{H}^s/mh(s)$ it follows that $\mathbb{E}(Y_1(s)) = 1$. Therefore $\mathbb{E}(Y_n(s)) = 1$ for all n's. The convergence of this family is the main result of the martingale theory [11].

(ii) The calculation of the variance will be done through the second moment of $\mathcal{H}_n^s$. By construction

$$\mathbb{E}\left((\mathcal{H}_{n+1}^s)^2 \,|\, \widehat{\mathcal{F}}_n\right) = \mathbb{E}\left(\sum_{u,u' \in \mathcal{V}_n} \kappa(u)^s \kappa(u')^s \sum_{v \in \mathrm{Ch}(u)} \sum_{v' \in \mathrm{Ch}(u')} \lambda_v^s \lambda_{v'}^s \,\Big|\, \widehat{\mathcal{F}}_n\right).$$

Let the terms with $u \neq u'$ be considered first. Then, since the λ_v's are independent for different v's, it follows that

$$\mathbb{E}\left(\kappa(u)^s\kappa(u')^s \sum_{v\in\mathrm{Ch}(u)}\sum_{v'\in\mathrm{Ch}(u')}\lambda_v^s\lambda_{v'}^s \,\Big|\,\widehat{\mathcal{F}}_n\right) = h(s)^2\mathbb{E}\left(\kappa(u)^s\kappa(u')^s\xi_u\xi_{u'}\,\Big|\,\widehat{\mathcal{F}}_n\right)$$

$$= h(s)^2 m^2 \kappa(u)^s\kappa(u')^s\,.$$

If now $u = u'$, this gives two terms: the first one are terms for which $v \neq v'$ and the other ones are for $v = v'$. The same type of calculation leads to

$$\mathbb{E}\left(\kappa(u)^{2s}\sum_{v\neq v'\in\mathrm{Ch}(u)}\lambda_v^s\lambda_{v'}^s\,\Big|\,\widehat{\mathcal{F}}_n\right) = h(s)^2(m^2+\sigma^2-m)\,\kappa(u)^{2s}\,,$$

$$\mathbb{E}\left(\kappa(u)^{2s}\sum_{v\in\mathrm{Ch}(u)}\lambda_v^{2s}\,\Big|\,\widehat{\mathcal{F}}_n\right) = h(2s)\,m\,\kappa(u)^{2s}\,.$$

Grouping these results, leads to

$$\mathbb{E}\left((\mathcal{H}_{n+1}^s)^2\,|\,\widehat{\mathcal{F}}_n\right) = h(s)^2\,m^2\,(\mathcal{H}_n^s)^2 + \{m\left(h(2s)-h(s)^2\right)+\sigma^2\,h(s)^2\}\mathcal{H}_n^{2s}\,.$$

Averaging on both sides gives

$$\mathrm{Var}(\mathcal{H}_{n+1}^s) = \{m\left(h(2s)-h(s)^2\right)+\sigma^2\,h(s)^2\}\,m^n h(2s)^n\,.$$

It is worth noticing that, thanks to the definition of h, the Cauchy-Schwarz inequality gives $h(s)^2 < h(2s)$. This inequality is actually strict because $\rho\{1\}\neq 1$, so that λ is not almost surely equal to one. Therefore

$$\mathrm{Var}(Y_n(s)) = \left(\frac{1}{m}\frac{h(2s)}{h(s)^2}\right)^n\left(1-(1-\frac{\sigma^2}{m})\frac{h(s)^2}{h(2s)}\right)\,.$$

It follows that, if $s < s_m$, then $h(s)m > 1$. In addition an elementary calculation shows that the map $g(s) = h(2s)/h(s)^2$ is monotone increasing, that $g(0) = 1$ and $\lim_{s\to\infty}g(s) = \rho\{1\}^{-1}$. Therefore, there is a unique $t_m > 0$ such that $m = g(t_m)$. Using the definition of s_m, it is easy to show that $s_m < t_m$. Hence

(a) if $s < t_m$, $\lim_{n\to\infty}\mathrm{Var}(Y_n(s)) = 0$, implying that Y_n converges almost surely to a constant; this constant can only be the common average, namely $\lim_{n\to\infty}Y_n(s) = 1$,

(b) if $s = t_m$, then the variance converges to a finite value

$$s = t_m \;\Rightarrow\; \lim_{n\to\infty}\mathrm{Var}(Y_n(t_m)) = \left(1+\frac{\sigma^2}{m^2}-\frac{1}{m}\right)\,,$$

(c) if $s > t_m$, then the limiting random variable $Y(s)$ does not have a finite second moment.

(iii) Let now $\Gamma' \subset \Gamma$ be two full finite subtrees of $\mathcal{T}$. Then there is a decreasing sequence of full finite subtrees such that $\Gamma' \subset \Gamma_j \subset \cdots \Gamma_1 \subset \Gamma_0 = \Gamma$, and such that Γ_{i+1} is obtained from Γ_i by the following procedure: each vertex $v \in \partial\Gamma_i$ which is not in $\partial\Gamma'$ is removed and replaced by its father. It is clear that, if Γ_i is full, so is Γ_{i+1}. This leads to

$$\mathbb{E}\left(\mathcal{H}^s(\Gamma)\,|\,\widehat{\mathcal{F}}_{\Gamma_1}\right) = \mathbb{E}\left(\sum_{u\in\partial\Gamma_1;\mathrm{Ch}(u)\cap\partial\Gamma'=\emptyset}\kappa(u)^s\sum_{v\in\mathrm{Ch}(u)}\lambda_v^s\,\Big|\,\widehat{\mathcal{F}}_{\Gamma_1}\right) + \sum_{u\in\partial\Gamma\cap\partial\Gamma'}\kappa(u)^s$$

Thanks to the definition of $\widehat{\mathcal{F}}_{\Gamma_1}$ the *r.h.s.* becomes

$$\mathbb{E}\left(\mathcal{H}^s(\Gamma)\,|\,\widehat{\mathcal{F}}_{\Gamma_1}\right) = mh(s)\sum_{u\in\partial\Gamma_1;\mathrm{Ch}(u)\cap\partial\Gamma'=\emptyset}\kappa(u)^s + \sum_{u\in\partial\Gamma\cap\partial\Gamma'}\kappa(u)^s\,.$$

In particular, if $s = s_m$, namely if $mh(s) = 1$, this gives $Y(\Gamma) = \mathcal{H}^{s_m}(\Gamma)$ so that

$$\mathbb{E}\left(Y(\Gamma)\,|\,\widehat{\mathcal{F}}_{\Gamma_1}\right) = Y(\Gamma_1)\,.$$

Proceeding inductively along the chain of Γ_i's, this gives

$$\mathbb{E}\left(Y(\Gamma)\,|\,\widehat{\mathcal{F}}_{\Gamma'}\right) = Y(\Gamma')\,.$$

Therefore the family $(Y(\Gamma))_{\Gamma\in\mathfrak{G}}$ is also a martingale *w.r.t.* the $\widehat{\mathcal{F}}_\Gamma$'s. The martingale theorem then shows that it converges almost surely. Since the full tree with boundary $\mathcal{V}_n$ is a member of this family and since it has been shown that the variance converges to zero (because $s_m < t_m$), the family converges to a constant almost surely.   □

Proposition 16. *Under the hypothesis of Proposition 15, the Hausdorff dimension of $(\partial\mathfrak{T}, d_\kappa)$ is almost surely equal to s_m.*

Proof: Thanks to Proposition 15, $\mathcal{H}^{s_m}(\Gamma)$ converges almost surely to 1. It follows that

$$\mathcal{H}_\delta^{s_m} = \inf_{\Gamma\in\mathfrak{G}_\delta}\mathcal{H}^{s_m}(\Gamma)\,,\qquad\Rightarrow\qquad\lim_{\delta\downarrow 0}\mathcal{H}_\delta^{s_m} = 1\,.$$

Consequently, $\mathcal{H}_\delta^s \to \infty$ for $s < s_m$ and $\mathcal{H}_\delta^s \to 0$ for $s > s_m$. Hence $\dim_H(\partial\mathfrak{T}, d_\kappa) = s_m$.   □

Proposition 17. *Under the hypothesis of Proposition 15, the Hausdorff measure of $(\partial\mathfrak{T}, d_\kappa)$ at the dimension $s = s_m$ exists almost surely and is a random probability.*

Proof: In order to prove it, it is sufficient to consider the basis of clopen sets of the form $[u]$ for $u \in \mathcal{V}$. It boils down to consider

$$\mathcal{H}^{s_m}(\Gamma; u) = \sum_{v\in\partial\Gamma;v\preceq u}\kappa(v)^s\,.$$

a calculation similar to the one made in the proof of Proposition 15, shows that the family of $\{\mathcal{H}^{s_m}(\Gamma; u)\,;\,\Gamma\in\mathcal{G}\,,\,u\in\mathcal{V}_\Gamma\}$ is also a martingale satisfying

$$\mathbb{E}\left(\mathcal{H}^{s_m}(\Gamma; u)\,|\,\widehat{\mathcal{F}}_{\Gamma_0}\right) = \kappa(u)^{s_m}\,,$$

for all full finite subtree Γ_0 with $u \in \partial\Gamma_0$ and $\Gamma \supset \Gamma_0$. Therefore, the martingale Theorem implies that $\mu([u]) = \lim_\Gamma \mathcal{H}^{s_m}(\Gamma; u)$ exists and that it is a random variable. Since the set of vertices is countable, the set of probability zero on which the convergence does not hold can be chosen independently on $u \in \mathcal{V}$. By construction

$$\sum_{u\in\partial\Gamma_0}\mathcal{H}^{s_m}(\Gamma; u) = \mathcal{H}^{s_m}(\Gamma)\,,$$

showing that, after taking the limit, $\sum_{u \in \partial \Gamma_0} \mu([u]) = 1$. Since this is true for any full finite subtree, the Proposition 6 shows that the same relation occurs if $\partial \Gamma_0$ is replaced by any finite partition of $\partial \mathcal{T}$ by clopen sets. Since clopen sets generates the σ-algebra of Borel sets, it shows that μ defines a probability measure on $\partial \mathcal{T}$. □

References

[1] J. Aliste, D. Coronel, "Tower systems for linearly repetitive Delone sets", *Ergod. Th. & Dynam. Sys.*, **31**, (2011),1595-1618.

[2] P. Arnoux, "Sturmian Sequences", Chap. 6 in N. Pytheas Fogg, *Substitutions in dynamics, arithmetics, and combinatorics*, Lecture Notes in Mathematics, Springer, (2002).

[3] K. B. Athreya, P.E. Ney, *Branching processes*, Die Grundlehren der mathematischen Wissenschaften, Band 196. Springer-Verlag, New York-Heidelberg, (1972). Reprinted by Dover Publications, Inc., Mineola, NY, (2004).

[4] J. Bellissard, R. Benedetti, J.-M. Gambaudo, "Spaces of tilings, finite telescopic approximations and gap-labeling", *Commun. Math. Phys.*, **261**, (2006), 1-41.

[5] J. Berstel, P. Séébold, "Sturmian Words", Chap.2 in M. Lothaire, *Algebraic Combinatorics on Words*, Cambridge University press, (2002).

[6] L. M. Blumenthal, "New theorems and methods in determinant theory", *Duke Mathematical Journal*, **2**, (1936), 396-404.

[7] L. Brouwer, "On the structure of perfect sets of points", *Proc. Akad. Amsterdam*, **12**, (1910), 785-794.

[8] P. Courrieu, "Straight monotonic embedding of data sets in Eucidean spaces", *Neural Networks*, *15*, (2002), 1185-1196.

[9] T. M. Cover, J. A. Thomas, *Elements of Information Theory*, Second Edition, Wiley, (2006).

[10] J. L. Doob, "Regularity properties of certain families of chance variables", *Trans. Amer. Math. Soc.*, **47**, (1940), 455-486.

[11] J. L. Doob, *Stochastic processes*, John Wiley & Sons, Inc., New York; Chapman & Hall, Limited, London, (1953).

[12] F. Durand, "Linearly recurrent subshifts have a finite number of non-periodic subshift factors", *Ergod. Th. & Dynam. Sys.* **20**, (2000), 1061-1078.

[13] F. Durand, "Corrigendum and addendum to 'Linearly recurrent subshifts have a finite number of non-periodic factors'", *Ergod. Th. & Dynam. Sys.* **23**, (2003), 663-669.

[14] F. Durand, B. Host, C. Skau, "Substitutions, Bratteli diagrams and dimension groups", *Ergod. Th. & Dynam. Sys.* **19**, (1999), 953-993.

[15] L. Euler, "De fractionibus continuis dissertation", *Comm. Acad. Sci. Petr.*, **9**, (1737), 98-137, 1744. Reprinted in *Leonhardi Euleri Opera Omnia*, Ser. I, Vol. **14**, Leipzig, Germany: Teubner, pp. 187-215, (1924).

[16] K. Falconer, *Fractal Geometry, Mathematical foundations and applications*, John Wiley & Sons, Ltd., Chichester, (1990).

[17] M. Fréchet, "Les dimensions d'un ensemble abstrait", *Mathematische Annalen*, **68**, (1910), 145168.

[18] T. Giordano, H. Matui, I. Putnam, C. Skau, "Orbit equivalence for Cantor minimal $\mathbb{Z}^d$-systems", *Inventiones Mathematicae*, *179* (2010), 119-158.

[19] T. E. Harris, *The theory of branching processes*, Die Grundlehren der mathematischen Wissenschaften, Bd. 119 Springer-Verlag, Berlin; Prentice-Hall, Inc., Englewood Cliffs, N.J. (1963). *The theory of branching processes* (Corrected reprint of the 1963 original book), Dover Phoenix Editions, Dover Publications, Inc., Mineola, NY, (2002).

[20] F. Hippert & D. Gratias Eds., *Lectures on Quasicrystals*, Editions de Physique, Les Ulis, (1994).

[21] P. Indyk, J. Matoušek, "Low-distorsion embeddings of finite metric spaces", in *Handbook of Discrete and Computational Geometry*, Chap. 8, p.177-196, Jacob E. Goodman, Joseph O'Rourke Eds, CRC Press, (2004).

[22] A. Julien, J. Savinien, "Embeddings of self-similar ultrametric Cantor sets", *Topology and Applications*, **158**, (2011), 2148-2157.

[23] D. G. Kendall, "Branching processes since 1873", *J. London Math. Soc.*, **41**, (1966), 385-406.

[24] A. Khintchine, *Continued Fractions*, New York: Dover, (1997).

[25] J. C. Lagarias, "Geometric models for quasicrystals I. Delone sets of finite type", *Discrete Comput. Geom.*, **21**, (1999), 161-191.

[26] P. Lévy, "Sur le développement en fraction continue d'un nombre choisi au hasard", (French) *Compositio Math.*, **3**, (1936), 286-303.

[27] C. A. Micchelli, "Interpolation of scattered data: Distance matrices and conditionally positive definite functions", *Constructive Approximation*, **2**, (1986), 11-22.

[28] G. Michon, "Les Cantors réguliers", *C. R. Acad. Sci. Paris Sér. I Math.*, **300**, (1985), 673-675.

[29] M. Morse, G. A. Hedlund, "Symbolic Dynamics II. Sturmian trajectories", *American Journal of Mathematics*, **62**, (1940), 1-42.

[30] J. Pearson, J. Bellissard, "Noncommutative Riemannian Geometry and Diffusion on Ultrametric Cantor Sets", *Journal of Noncommutative Geometry*, **3**, (2009), 447-480.

[31] Sadun L., *Topology of tiling spaces*. University Lecture Series, **46**. American Mathematical Society, Providence, RI, 2008.

[32] I. J. Schoenberg, "Remarks to Maurice Fréchet's article "Sur la définition axiomatique d'une classe d'espace distanciés vectoriellement applicable sur l'espace de Hilbert", *Ann. of Math.*, **36**, (1935), 724732.

[33] I. J. Schoenberg, "On certain metric spaces arising from euclidean spaces by a change of metric and their imbedding in Hilbert space", *Annals of Mathematics*, **38**(2), (1937), 787-793.

[34] I. J. Schoenberg, "Metric spaces and positive definite functions", *Transactions of the American Mathematical Society*, **44**, (1938), 522-536.

[35] H. W. Watson, F. Galton, "On the Probability of the Extinction of Families", *Journal of the Anthropological Institute of Great Britain*, **4**, (1875), 138-144.

Jean V. Bellissard, Georgia Institute of Technology, School of Mathematics, Atlanta GA 30332-0160

E-mail address: jeanbel@math.gatech.edu

Antoine Julien, Department of Mathematics and Statistics, University of Victoria, Victoria, BC, Canada V8W 3R4

E-mail address: antoinej@uvic.ca