# Covering Space in the Besicovitch Topology*

Julien Cervelle

LACL, Université Paris-Est Créteil
94010 Créteil cedex, France
julien.cervelle@univ-paris-est.fr

**Abstract.** This paper studies how one can spread points in the Besicovitch space in order to keep them far one from another. We first study the general case and then what happens if the chosen points are all regular Toeplitz configurations or all quasiperiodic configurations.

**Keywords:** Hamming distance, Besicovitch distance, dynamical systems, Toeplitz sequences.

## 1 Introduction

In compact spaces, the more you have points in a set, the shorter the distance between the closest points of the set. More precisely, for any $\varepsilon$, there is an integer $N$ such that any set of cardinal at least $N$ contains two points whose relative distance is less than $\varepsilon$. This is easily proved covering the compact space with open balls of diameter $\varepsilon$ and selecting a finite sub-covering and choosing $N$ as its size plus one. If one has $N$ points, using pigeon hole lemma, there are two points in the same ball.

This is not the case in non-compact spaces. For instance, it is clear that $\mathbb{N} \subset \mathbb{R}$ is an infinite set of reals which are all at distance greater or equal to 1 from all other members.

This feature has some direct consequences on code theory. For error-correcting (resp. detecting) codes, the valid representations of information must be at the center of open balls of some fixed radius which do not overlap (resp. which do not contain another valid representation). The radius depends on the number of error to correct (resp. detect). For classical code theory on finite words, the Hamming distance on words is often considered.

In this paper, we study the space $\{0,1\}^{\mathbb{N}}$ of uni-infinite words on $\{0,1\}$ or *configurations* endowed with the Besicovitch topology. Configurations are often called sequences or streams. For more investigations about configurations, see for instance [1]. The Besicovitch distance is used, among others, in the domain of symbolic dynamical systems, and particularly cellular automata. Considering the phase space $\{0,1\}^{\mathbb{N}}$, the classical product topology, called Cantor topology, has counter-intuitive properties and the Besicovitch topology was proposed as

* This work has been supported by the ANR Blanc “Projet EQINOCS” (ANR-11-BS02-004).

A.-H. Dediu and C. Martín-Vide (Eds.): LATA 2012, LNCS 7183, pp. 169–178, 2012.

a less biased alternative for cellular automata dynamical behavior study (see [4,3,6]). Moreover, looking at its definition, one can note that it is a sort of extension of the Hamming distance to infinite words.

The space $\{0,1\}^{\mathbb{N}}$ is not compact and many of the proofs for result with Cantor topology can't be done in the Besicovitch topology. For instance Hedlund's theorem which states that cellular automata are the continuous shift-invariant map on $\{0,1\}^{\mathbb{N}}$ is based on compactness and open cover extraction (see [7]).

We want to evaluate how much non-compact the Besicovitch space is and, if possible, prove a kind of weak compactness. We do this by studying how points get closer as you add points. Formally if $\mathcal{S}$ is a set with at least $N$ members we want to find the maximum distance between the closest members of $\mathcal{S}$ and see how this behaves when $N$ tends to infinity.

First, we state a negative result giving an uncountable set of configurations such that any two members are at distance 1, the maximum for the Besicovitch distance. Hence, there is no chance to prove a kind of weak compactness for the Besicovitch space. Next, we try to see if the negative result is still true restricting ourselves to some natural subsets of configurations, *Toeplitz configurations* and *quasiperiodic configurations*.

Toeplitz configuration are a dense positively invariant set in $\{0,1\}^{\mathbb{N}}$. It has been studied in [2] and proposed as a good test set since to prove some properties it is enough to prove them only on Toeplitz sequences. They play a role similar to periodic sequences in Cantor topology. Quasiperiodic configurations are a natural candidate between the general case and Toeplitz configurations and play a special role in the field of tilings, often considered as the static version of cellular automata (see for instance [5]).

We first prove that the negative result still holds on quasiperiodic configurations. However, we can prove a non-trivial bound for a natural subset of Toeplitz configurations: *regular* Toeplitz configurations. In order to perform this last study, we first consider finite words on $\{0,1\}$ of a given length since the Besicovitch distance definition is expressed in terms of the Hamming distance between the prefixes. Then we extend the result on the Hamming distance to $\{0,1\}^{\mathbb{N}}$ proving that the distance between closest members of a set tends to one half when the cardinal of the set increases and that this bound is tight.

The paper is organized as follows. In the next section, we give definitions about the Besicovitch distance, quasiperiodic and Toeplitz configurations. In Section 3 and 4 we state the negative results about the Besicovitch distance for the general and the quasiperiodic cases. In Section 5, we study the restriction to regular Toeplitz configurations.

## 2 Definitions and Tools

In this section, we introduce the space we study and a few notions.

*Configurations.* We call *configurations* uni-infinite words on $\{0,1\}$. If $x$ is a configuration, as for words, we note $x_i$ the $i$th letter (the first one has index 0). We note $x_{\rightarrow n}$ the prefix on length $n$ of $x$ i.e. $x_{\rightarrow n} = x_0 \ldots x_{n-1}$. A word $u$ is a *factor* of the configuration (or the word) $x$ if there is an integer $k$ such that for all $i$ where $0 \leqslant i < |u|$, $u_i = x_{k+i}$. We note it $u \sqsubset x$ or $u \sqsubset_n x$ if $u$ is of length $n$.

*The Besicovitch topology.* The Besicovitch topology measures the rate of differences between two configurations. Its formal definition is given by the pseudo-distance $d_B$ defined by

$$d_B(x, y) = \limsup_{n\to\infty} \frac{d_H(x_{\to n}, y_{\to n})}{n}$$

where $d_H$, the Hamming distance, is such that $d_H(x, y) = |\{n,\ x_n \neq y_n\}|$.

It is only a pseudo-distance since two configurations with finitely or logarithmically many differences are at distance zero. Taking the quotient of $\{0, 1\}^{\mathbb{N}}$ w.r.t. the equivalence relation $x \sim y \Leftrightarrow d(x, y) = 0$, we obtain a distance. For more information about the Besicovitch topology, see [4,3].

*Quasiperiodic configurations.* The configuration $x$ is *quasiperiodic* if

$$\forall n, \exists N, \forall u, u \sqsubset_n x \Rightarrow \forall w, w \sqsubset_N x \Rightarrow u \sqsubset w,$$

that is if for all integers $n$ there exists an integer $N$ such that all factor of length $n$ of $x$ can be found in any factor of length $N$.

### 2.1 Toeplitz Configurations and Their Construction

A configuration $x$ is *Toeplitz* if for all positions $i \in \mathbb{N}$, there exists a period $p$ such that $\forall k \in \mathbb{Z}$ such that $pk + i \geqslant 0$, $x_{pk+i} = x_i$.

In order to build Toeplitz configurations, one can use a simple algorithm. It assigns letters to cells of the configuration step by step. Initially, all cells are unassigned. At each step, a cell is assigned a value, and this assignment is repeated periodically along the configuration. With this algorithm, one can build all Toeplitz configurations.

Formally, a Toeplitz configuration $x$ is totally (but not uniquely) defined by a finite or infinite sequence of couple $(v_i, p_i)_{0\leqslant i<L}$ ($L \in \mathbb{N} \cup \{\infty\}$) where the cells of $x$ are filled with values $v_i$ periodically with period $p_i$. At step 0, $x_0$ and all cells of index $kp_0$ for $k \in \mathbb{N}$ are assigned to $v_0$. At step 1, choose the smallest index of an unassigned cell $j_1$ (which must be 1 unless $p_0 = 1$ in which case there is no step 1). All cells of index $j_1 + kp_1$ are assigned to $v_1$. We continue to assign values to cells periodically, always starting from the unassigned cell of smallest index: at step $i$, for $j_i$ the smallest index of an unassigned cell, all cells of index $j_i + kp_i$ are assigned to $v_i$. The beginning of the process is illustrated Figure 1

This process must verify two conditions:

- the periods $p_i$ must be chosen so that a cell is never assigned twice. This implies a rather complicated relation on periods;
- each cell is eventually assigned at some step.

If $v$ and $p$ are two such sequences, we note $\mathcal{A}(v, p)$ the Toeplitz sequence the algorithm outputs. If the sequence $(v_i, p_i)_{i\in\mathbb{N}}$ is finite, then the configuration is periodic with period given by the least common multiple of the $p_i$.

$$
\begin{array}{rl}
p_0 = 3: & \mathbf{v_0}\ .\ .\ \mathbf{v_0}\ .\ .\ \mathbf{v_0}\ .\ .\ \mathbf{v_0}\ .\ .\ \mathbf{v_0}\ .\ .\ \mathbf{v_0}\ .\ .\ \mathbf{v_0}\ .\ .\ \mathbf{v_0}\ .\ . \\
p_1 = 6: & v_0\,\mathbf{v_1}\ .\ v_0\ .\ .\ v_0\,\mathbf{v_1}\ .\ v_0\ .\ .\ v_0\,\mathbf{v_1}\ .\ v_0\ .\ .\ v_0\,\mathbf{v_1}\ .\ v_0\ .\ . \\
p_2 = 6: & v_0\,v_1\,\mathbf{v_2}\,v_0\ .\ .\ v_0\,v_1\,\mathbf{v_2}\,v_0\ .\ .\ v_0\,v_1\,\mathbf{v_2}\,v_0\ .\ .\ v_0\,v_1\,\mathbf{v_2}\,v_0\ .\ . \\
p_3 = 12: & v_0\,v_1\,v_2\,v_0\,\mathbf{v_3}\ .\ v_0\,v_1\,v_2\,v_0\ .\ .\ v_0\,v_1\,v_2\,v_0\,\mathbf{v_3}\ .\ v_0\,v_1\,v_2\,v_0\ .\ . \\
p_4 = 6: & v_0\,v_1\,v_2\,v_0\,v_3\,\mathbf{v_4}\,v_0\,v_1\,v_2\,v_0\ .\ \mathbf{v_4}\,v_0\,v_1\,v_2\,v_0\,v_3\,\mathbf{v_4}\,v_0\,v_1\,v_2\,v_0\ .\ \mathbf{v_4} \\
 & \vdots \\
 & v_0\,v_1\,v_2\,v_0\,v_3\,v_4\,v_0\,v_1\,v_2\,v_0\,v_5\,v_4\,v_0\,v_1\,v_2\,v_0\,v_3\,v_4\,v_0\,v_1\,v_2\,v_0\,v_6\,v_4
\end{array}
$$

**Fig. 1.** Sample Toeplitz configuration construction

## 3 The Result for the General Case

Our first result for the Besicovitch topology is negative. It states that there exists an uncountable set of configurations such that each member is at distance 1 (the maximum for the Besicovitch distance) from all other members.

**Lemma 1.** *There exists an uncountable set $\mathcal{D}$ of infinite words such that for any two of those words, they differ at infinitely many positions.*

*Proof.* Let $u$ be a sequence of $\{0,1\}^{\mathbb{N}}$. Let $w^u$ be the Toeplitz configurations built on the sequence $(u_i, 2^i)_{i\in\mathbb{N}}$.

Informally, the sequence is built as follows: the half of the configuration is assigned to $u_0$; the half of the remaining cells is assigned to $u_1$; the half of the remaining cells is assigned to $u_2$ and so on. The beginning of the sequence is (letters are raised according to their index for better readability)

$$u_0{}^{u_1}u_0{}^{u_2}u_0{}^{u_1}u_0{}^{u_3}u_0{}^{u_1}u_0{}^{u_2}u_0{}^{u_1}u_0{}^{u_4}u_0{}^{u_1}u_0{}^{u_2}u_0{}^{u_1}u_0{}^{u_3}u_0{}^{u_1}u_0{}^{u_2}u_0{}^{u_1}u_0$$

If $u$ and $v$ are two distinct binary sequences and $l$ is such that $u_l \neq v_l$, then at all positions $i$ such that $i \equiv 2^l - 1 \mod 2^{l+1}$ (hence infinitely many), the sequences $w^u$ and $w^v$ are distinct. The set $\mathcal{D} = \{w^u,\, u \in \{0,1\}^{\mathbb{N}}\}$ proves the lemma. □

The negative result is as follows.

**Proposition 2.** *There exists an uncountable set of configurations $\mathcal{Y}$ such that for any $x$ and $y$ in $\mathcal{Y}$, $d_B(x,y) = 1$.*

*Proof.* In this proof, we note $\mathfrak{2}^i = 2^{2^i}$. Let $w$ be a sequence of $\mathcal{D}$ of Lemma 1. Define the configuration $s^w$ by $s^w_i = w_j$ for $\mathfrak{2}^j - 2 \leqslant i < \mathfrak{2}^{j+1} - 2$. It is the concatenation of blocs of $\mathfrak{2}^{j+1} - \mathfrak{2}^j$ times the letter $w^j$.

Let $w$ and $w'$ be two distinct sequences of $\mathcal{D}$. There is an infinite set $J$ of positions which $w$ and $w'$ differ at. By definition of $s$, for all $j \in J$, since $\mathfrak{2}^i = o(\mathfrak{2}^{i+1})$, one has

$$\frac{d_H(s^w_{\to\mathfrak{2}^{j+1}-2}, s^{w'}_{\to\mathfrak{2}^{j+1}-2})}{\mathfrak{2}^{j+1}-2} \geqslant \frac{(\mathfrak{2}^{j+1}-2)-(\mathfrak{2}^j-2)}{\mathfrak{2}^{j+1}-2} \sim 1 \ .$$

Hence, considering that the limit superior is greater than the limit superior of any subsequence, $d_B(s^w, s^{w'}) = 1$. The set $\mathcal{Y}$ is defined by $\mathcal{Y} = \{s^w,\, w \in \mathcal{D}\}$. □

The following corollary can be deduced:

**Corollary 3.** *Let $\mathcal{R}$ be a dense set of configurations according to Besicovitch topology. Then for all $\varepsilon > 0$, there is an infinite set $\mathcal{E}_{\mathcal{R}}^{\varepsilon}$ of configurations in $\mathcal{R}$ such that for any $x$ and $y$ in $\mathcal{E}_{\mathcal{R}}^{\varepsilon}$, $d_B(x, y) > 1 - \varepsilon$. If $\mathcal{R}$ is uncountable, $\mathcal{E}_{\mathcal{R}}^{\varepsilon}$ can be chosen uncountable.*

*Proof.* Let $\mathcal{Y}$ be the set of proposition 2. As $\mathcal{R}$ is dense, for all $y \in \mathcal{Y}$, there is a point $r_y \in \mathcal{R}$ such that $d_B(y, r_y) < \frac{\varepsilon}{2}$. Denote $\mathcal{E}_{\mathcal{R}}^{\varepsilon} = \{r_y,\, y \in \mathcal{Y}\}$. Provided $\varepsilon$ is less than 1, $\mathcal{Y}$ and $\mathcal{E}_{\mathcal{R}}^{\varepsilon}$ have same cardinal. Moreover, using triangular inequality, one has that for all $x$ and $y$ in $\mathcal{Y}$, $d_B(r_x, r_y) > d_B(x, y) - d_B(r_x, x) - d_B(r_y, y) > 1 - \varepsilon$. □

## 4 Quasiperiodic Case Study

In this section, we show that the negative result of the previous section still holds for quasiperiodic configurations.

**Theorem 4.** *There exists a non countable set $\mathcal{Q}$ of quasiperiodic configurations such that for any two $x$ and $y$ in $\mathcal{Q}$, $d_B(x, y) = 1$.*

*Proof.* In this proof, we note $\overline{u}$ the word $u$ where all ones are replaced by zeros and zeros by ones. Let $u \in \{0, 1\}^{\mathbb{N}}$. Let $q_u$ defined as the limit of the substitution process: $a_0 = 0$; $a_{n+1} = a_n^2\ \overline{a_n}\ ^2 a_n\ \overline{a_n}\ ^n$ if $u_n = 0$ and $a_{n+1} = a_n^2\ \overline{a_n}\ ^2 a_n a_n^n$ if $u_n = 1$.

Let us first prove that any sequence $q_u$ is quasiperiodic. Let $w$ be a pattern of size $\ell$ of $q_u$. Let $n$ be such that $\ell < |a_n|$. By construction, $w$ is a factor of either $a_n a_n$, $a_n\ \overline{a_n}$, $\overline{a_n a_n}$ or $\overline{a_n}\ a_n$. By construction, in each window of size $2|a_{n+1}|$ either $a_{n+1}$ of $\overline{a_{n+1}}$ occurs. Both contains each of the words $a_n a_n$, $a_n\ \overline{a_n}$, $\overline{a_n a_n}$ or $\overline{a_n}\ a_n$ and so each window of size $2|a_{n+1}|$ contains $w$.

Define $\mathcal{Q} = \{q_u, u \in \mathcal{D}\}$, where $\mathcal{D}$ is defined in lemma 1. Let $u, v \in \mathcal{D}$. Let $I$ be the infinite set of positions $i$ such that $u_i \neq v_i$. We have that

$$d_B(q_u, q_v) \geqslant \lim_{n \in I} \frac{d_H(a_n^2\ \overline{a_n}\ ^2 a_n\ \overline{a_n}\ ^n, a_n^2\ \overline{a_n}\ ^2 a_n a_n^n)}{|a_{n+1}|} \geqslant \lim_{n \in I} \frac{n|a_n|}{(5+n)|a_n|} = 1\ .$$

□

## 5 Regular Toeplitz Case Study

In this section, we deal with regular Toeplitz sequences that will be defined in Section 5.2.

### 5.1 Finite Words and the Hamming Distance

In this section, we study the space $\{0, 1\}^{\ell}$ of words on alphabet $\{0, 1\}$ of a given length $\ell$, endowed with the Hamming distance $d_H$. For a set of cardinal $N$, we want to find a (tight) bound $M$ such that $\forall \mathcal{S} \subset \{0, 1\}^{\ell}$ s.t. $|\mathcal{S}| = N$, $\min\{d_H(x, y) |\, x, y \in \mathcal{S}\} \leqslant M$.

**Some Technical Lemmas.** We give two combinatorial lemmas.

We consider a set $\mathcal{S} = \{w^0, \dots, w^{N-1}\}$ of $N$ words or length $\ell$. For $i \in \{0, \dots, \ell-1\}$, we define $\mathcal{P}_i = \{\{p,q\} | w_i^p = w_i^q\}$. For a couple $\{p,q\}$, $d_h(w^p, w^q) = \ell - m$, where $m = |\{i, \{p,q\} \in \mathcal{P}_i\}|$. The minimum Hamming distance between words of $\mathcal{S}$ is linked to the couple which maximizes $m$.

In order to simplify proofs and results, we use the following function,

$$g(n) = \binom{\lfloor \frac{n}{2} \rfloor}{2} + \binom{\lceil \frac{n}{2} \rceil}{2} \quad \text{for } n \geqslant 4 \ .$$

The first lemma considers a vector $V = (V_0, \dots, V_{N-1})$ and bounds the number of couples $\{p,q\}$ such that $V_p = V_q$.

**Lemma 5.** *Let $V$ be a vector of letters from the alphabet $\{0,1\}$ of size $N \geqslant 4$. We have that $\left|\{\{p,q\},\, V_p = V_q\}\right| \geqslant g(N)$.*

*Proof.* Let $z$ be the number of $i$ such that $V_i = 0$. There are $\binom{z}{2}$ couples $\{p,q\}$ such that $V_p = V_q = 0$ and $\binom{N-z}{2}$ couples $\{p,q\}$ such that $V_p = V_q = 1$. Hence $|\{\{p,q\},\, V_p = V_q\}| = \binom{N-z}{2} + \binom{z}{2}$. The proof is achieved by a straightforward recurrence, proving that for all $n \geqslant 4$ and for all $z$ such that $0 \leqslant z \leqslant \frac{n}{2}$, $\binom{n-z}{2} + \binom{z}{2} \geqslant g(n)$. □

The next lemma is an upper bound for $M$, the number of $\mathcal{P}_i$ which contains the couple that occurs the most often.

**Lemma 6.** *Let $\mathcal{I}$ be a set of finite cardinal $c$. Let $(\mathcal{P}_i)_{0 \leqslant i < \ell}$ be a sequence of subsets of $\mathcal{I}$ $|\mathcal{P}_i| \geqslant k$. One has $\max_{s \in \mathcal{I}} |\{i,\, s \in \mathcal{P}_i\}| \geqslant \frac{\ell k}{c}$.*

*Proof.* Let $M = \max_{s \in \mathcal{I}} |\{i,\, s \in \mathcal{P}_i\}|$. One has

$$\ell k \leqslant \sum_{i=0}^{\ell-1} |\mathcal{P}_i| = \sum_{s \in \mathcal{I}} \sum_{i=0}^{\ell-1} |\mathcal{P}_i \cap \{s\}| = \sum_{s \in \mathcal{I}} |\{i,\, s \in \mathcal{P}_i\}| \leqslant Mc \ .\square$$

**Bound for the Hamming Distance.** Let $f$ be defined by

$$f(n) = \frac{\lceil \frac{n}{2} \rceil}{2 \lceil \frac{n}{2} \rceil - 1} \ .$$

Note that $f(n) = 1 - \frac{g(n)}{\binom{n}{2}}$, $\lim_{n \to \infty} f(n) = \frac{1}{2}$ and $f(2n) = f(2n-1)$.

**Proposition 7.** *Let $\mathcal{S}$ be a set of $N$ words of length $\ell$ from alphabet $\{0,1\}$. Then $\min_{x,y \in \mathcal{S}} d_H(x,y) \leqslant f(N)\ell$.*

The result states that for a set of cardinal $N$ of words of fixed length, one can always find two words whose ratio of differences relatively to their length is less than $f(N)$, which tends to one half when $N$ is large enough.

*Proof.* Let $\mathcal{S} = \{w^0, \dots, w^{N-1}\}$. For all $i \in \{0, \dots, \ell-1\}$, let $\mathcal{P}_i = \big\{\{p,q\},\ w_i^p = w_i^q\big\}$ and $M = \max_c |\{i, c \in \mathcal{P}_i\}|$.

Let $w^p$ and $w^q$ be two words. One has $\forall i, \{p,q\} \in \mathcal{P}_i \Rightarrow w_i^p = w_i^q$. Hence $d_H(w^p, w^q) = \ell - |\{i,\ \{p,q\} \in \mathcal{P}_i\}| \geqslant \ell - M$, and equality holds for the couple $\{p,q\}$ such that $|\{i, \{p,q\} \in \mathcal{P}_i\}| = M$. We conclude that $\ell - M$ is the distance between the closest words $w_p$ and $w_q$. By Lemma 5 we have that $|\mathcal{P}_i| \geqslant g(N)$. Using this inequality in Lemma 6 one finds that $M = \max_c |\{i, c \in \mathcal{P}_i\}| \geqslant \frac{\ell g(N)}{\binom{N}{2}} = \ell(1 - f(N))$ and that $d_H(w_p, w_q) = \ell - M \leqslant \ell - \ell(1 - f(N)) = f(N)\ell$. □

**Bound Tightness.** These kind of bounds have already been studied in the field of code theory in [8]. This paper studies equidistant codes and gives a result which states the tightness of bound of Proposition 7. For all $N$, it gives a set $\mathcal{W}_N$ of $N$ words of some fixed length $\ell$ at Hamming distance $f(N)\ell$ one another. Words of $\mathcal{W}_N$ for $N = 4$, 6, 8 are given in Figure 2, where $\mathcal{C}$ is ordered lexicographically.

| | | |
|---|---|---|
| | | $w^0 =$ 11111111111111111111111111111111111 |
| | $w^0 =$ 1111111111 | $w^1 =$ 11111111111111100000000000000000000 |
| $w^0 =$ 111 | $w^1 =$ 1111000000 | $w^2 =$ 11111000000000011111111110000000000 |
| $w^1 =$ 100 | $w^2 =$ 1000111000 | $w^3 =$ 10000111100000011110000001111110000 |
| $w^2 =$ 010 | $w^3 =$ 0100100110 | $w^4 =$ 01000100011100010001110001110001110 |
| $w^3 =$ 001 | $w^4 =$ 0010010101 | $w^5 =$ 00100010010011001001001101001101101 |
| | $w^5 =$ 0001001011 | $w^6 =$ 00010001001010100100101010101011011 |
| | | $w^7 =$ 00001000100101100010010110010110111 |
| *(a)* $2n = 4$ | *(b)* $2n = 6$ | *(c)* $2n = 8$ |

```
111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111
111111111111111111111111111111111111111111111111111111110000000000000000000000000000000000000000000000000000000000000000000000
111111111111111111111000000000000000000000000000000000001111111111111111111111111111111111100000000000000000000000000000000000
111111000000000000000111111111111111000000000000000000001111111111111110000000000000000000011111111111111111111000000000000000
100000111110000000000111110000000000111111111100000000001111100000000001111111111000000000011111111110000000000111111111100000
010000100001111000000100001111000000111100000011111100001000011110000001111000000111111000011110000001111110000111111000011110
001000010001000111000010001000111000100011100011100011100100010001110001000111000111000111010001110001110001110111000111011101
000100001000100100110001000100100110010010011010011011010010001001001100100100110100110110101001001101001101101100110110111011
000010000100010010101000100010010101001001010101010110110001000100101010010010101010101101100100101010101011011010101101110111
000001000010001001011000010001001011000100101100101101110000100010010110001001011001011011100010010110010110111001011011101111
```

*(d)* $2n = 10$

**Fig. 2.** Words for small $n$

## 5.2 Result for Regular Toeplitz Configurations

The obstacle to the extension of the result on finite words to infinite words comes from the limit superior in the definition of $d_B$. If we restrict our consideration to a class of configurations on which the limit of the Hamming distances between prefixes necessarily exists and is therefore the limit superior required in the Besicovitch distance, then the extension is possible.

In this section, we prove that the limit exists for subset of Toeplitz configurations called *regular Toeplitz configurations*. From corollary 3, as Toeplitz configurations are dense, there is no hope to extend it to the whole class.

During the building of a Toeplitz confiugration using the algorithm given in Section 2.1, though all cells are to be assigned, there are no insurance that the proportion of assigned cells tends to 1 while going through steps. We are only sure that each cell is eventually defined at some step. However, as Besicovitch topology relies on proportions, we say that a Toeplitz configuration $x$ is *regular* if there are sequences $p$ and $v$ such that $x = \mathcal{A}(v, p)$ and:

$$\sum_{0 \leqslant j < L} \frac{1}{p_j} = 1 \ . \tag{$*$}$$

The set of regular Toeplitz sequences has continuous cardinal since the sequence $p_i = 2^{i+1}$ generates a regular Toeplitz sequence for any $v$.

We prove that when the cardinal of a set of regular Toeplitz configurations increases, the closest members tend to be at distance one half from each-other.

First we have to state that the limit exists.

**Lemma 8.** *If $x$ and $y$ are regular Toeplitz configurations, the sequence $u_n = \frac{d_H(x_{\to n}, y_{\to n})}{n}$ is a Cauchy sequence.*

*Proof.* Let $(v_i, p_i)_{0 \leqslant i < L}$ be a sequence defining $x$ and $(v'_i, p'_i)_{0 \leqslant i < L'}$ be a sequence defining $y$ using the algorithm given in Section 2.1.

Let $\varepsilon > 0$. Using Equation $(*)$ there is an integer $N$ such that

$$\sum_{j=0}^{N} \frac{1}{p_j} \geqslant 1 - \frac{\varepsilon}{8} \quad \text{and} \quad \sum_{j=0}^{N} \frac{1}{p'_j} \geqslant 1 - \frac{\varepsilon}{8} \ .$$

Then at step $N$ in building $x$ and $y$, a proportion $1 - \frac{\varepsilon}{8}$ cells has been assigned. Moreover, all the letters in these assigned cells repeat with a period $P = \operatorname{lcm}\{p_i, p'_i\, i \in \{0, \dots, N\}\}$. Hence, there are two words $M_x$ and $M_y$ of length $P$ on alphabet $\{0, 1, \#\}$, where $x$ [resp. $y$] is the repetition of $M_x$ [resp. $M_y$] where # can be replaced by 0 or 1. As a ratio of $1 - \frac{\varepsilon}{8}$ positions are already set, $M_x$ and $M_y$ each has at most $P\frac{\varepsilon}{8}$ occurrences of #.

For instance, if $N = 2$, $v_0 = 0$, $v'_0 = 1$, $v_1 = 1$, $v'_1 = 0$, $p_0 = p'_0 = 2$, $p_1 = 4$ and $p'_1 = 6$ then

$$x = 010?010?010?010?010?010?010?010?010?010?010?010?\dots$$
$$y = 101?1?101?1?101?1?101?1?101?1?101?1?101?1?101?1?\dots$$

where cells marked with ? is set by other values of $v$, $v'$, $p$ and $p'$. The repeated words are $M_x = 010\#010\#010\#$ and $M_y = 101\#1\#101\#1\#$.

If $d = d_H(M_x, M_y)$, for $n > P$ we have $d \left\lfloor \frac{n}{P} \right\rfloor \leqslant d_H(x_{\to n}, y_{\to n}) \leqslant \left(d + \frac{P\varepsilon}{8}\right) \left\lceil \frac{n}{P} \right\rceil$ and hence $\frac{d\left(\frac{n}{P} - 1\right)}{n} \leqslant \frac{d_H(x_{\to n}, y_{\to n})}{n} \leqslant \frac{\left(d + \frac{P\varepsilon}{8}\right)\left(\frac{n}{P} + 1\right)}{n}$.

We conclude that $\left|u_n - \frac{d}{P}\right| = \left|\frac{d_H(x_{\to n}, y_{\to n})}{n} - \frac{d}{P}\right| \leqslant \frac{\varepsilon}{8P} + \frac{d}{n} + \frac{P\varepsilon}{8n} \leqslant \frac{\varepsilon}{4} + \frac{d}{n}$.

For $n \geqslant \max\left(P, \frac{4d}{\varepsilon}\right) = N'$, one has $\left|u_n - \frac{d}{P}\right| \leqslant \frac{\varepsilon}{2}$. Using the triangular inequality, one has, for all $n$ and $m$ greater than $N'$, $|u_n - u_m| \leqslant \left|u_n - \frac{d}{P}\right| + \left|\frac{d}{P} - u_m\right| \leqslant \varepsilon$.

We conclude that $u_n$ is a Cauchy sequence. □

Now we can state the proposition for regular Toeplitz configurations.

**Proposition 9.** *If $S$ is a set of $N$ regular Toeplitz configurations,*

$$\min_{x,y\in\mathcal{S}} d_B(x,y) \leqslant f(N)$$

*where $f$ is the function defined in Section 5.1.*

*Proof.* Using Lemma 8 and Proposition 7, we have:

$$\begin{aligned}\min_{x,y\in\mathcal{S}} d_B(x,y) &= \min_{x,y\in\mathcal{S}} \limsup_{n\to\infty} \frac{d_H(x_{\to n}, y_{\to n})}{n} \\ &= \min_{x,y\in\mathcal{S}} \lim_{n\to\infty} \frac{d_H(x_{\to n}, y_{\to n})}{n} \leqslant f(N) \ . \end{aligned}$$ □

The following corollary applies the above result to an infinite set of Toeplitz configurations.

**Corollary 10.** *If $\mathcal{S}$ is an infinite set of regular Toeplitz configurations, then for all $\varepsilon > 0$, one can find infinitely many pairs $(x,y)$ with $x$ and $y$ in $\mathcal{S}$ such that*

$$d_B(x,y) \leqslant \frac{1}{2} + \varepsilon$$

*Proof.* Let $N$ be such that $f(N) \leqslant \frac{1}{2}+\varepsilon$. Picking $N$ elements of $\mathcal{S}$, one can apply Proposition 9 to get two elements $x$ and $y$ at distance less than $f(N) \leqslant \frac{1}{2}+\varepsilon$. Picking $N$ other elements, one can get two more elements $x$ and $y$ verifying the condition. Repeating this process, one can find infinitely many $x$ and $y$. □

### 5.3 Bound Tightness for Regular Toeplitz Configurations

As the result on regular Toeplitz configurations comes from Proposition 7 on finite words with the Hamming distance, it can be adapted to configurations.

First, we need a simple lemma giving the relation between the Hamming distance and the Besicovitch distance for periodic configurations.

**Lemma 11.** *Let $u$ and $v$ be finite words of same length $\ell$, $x = u^\infty$ and $y = v^\infty$ the periodic configurations whose repeated pattern are $u$ and $v$ respectively. Then $d_B(x,y) = \frac{d_H(u,v)}{\ell}$.*

*Proof.* As $x_{\to n} = u^{\lfloor \frac{n}{\ell} \rfloor} u_{\to n - \lfloor \frac{n}{\ell} \rfloor}$ and $y_{\to n} = v^{\lfloor \frac{n}{\ell} \rfloor} v_{\to n - \lfloor \frac{n}{\ell} \rfloor}$, one has

$$\begin{aligned} d_B(x,y) &= \limsup_{n\to\infty} \frac{d_H(x_{\to n}, y_{\to n})}{n} \\ &= \limsup_{n\to\infty} \frac{1}{n}\left( \left\lfloor \frac{n}{\ell} \right\rfloor d_H(u,v) + d_H(u_{\to n - \lfloor \frac{n}{\ell} \rfloor}, v_{\to n - \lfloor \frac{n}{\ell} \rfloor}) \right) \\ &= \limsup_{n\to\infty} \frac{d_H(u,v)}{n} \left\lfloor \frac{n}{\ell} \right\rfloor + \frac{d_H(u_{\to n - \lfloor \frac{n}{\ell} \rfloor}, v_{\to n - \lfloor \frac{n}{\ell} \rfloor})}{n} \ . \end{aligned}$$

The first term of the sum tends to $\frac{d_H(u,v)}{\ell}$ and the second to 0. □

The following result gives bound tightness.

**Proposition 12.** *For all integer $N$, there is a set $\mathcal{X}_N$ of cardinal $N$ of periodic (hence regular Toeplitz) configurations such that*

$$\forall u, v \in \mathcal{X}_N, d_B(u,v) = f(N) \ .$$

*There is an infinite set $\mathcal{X}_\infty$ of periodic configurations such that*

$$\forall u, v \in \mathcal{X}_\infty, d_B(u,v) = \frac{1}{2} \ .$$

*Proof.* Let $\mathcal{W}_N$ be the set introduced at the end of Section 5.1. Let $\mathcal{X}_N$ be the set of periodic configurations whose repeated words are the words of $\mathcal{W}_N$. Using Lemma 11, since $\forall x, y \in \mathcal{W}_N$, $d_H(x,y) = |x|f(N)$, one has $\forall u, v \in \mathcal{C}, d_B(u,v) = f(N)$.

Let $\mathcal{X}_\infty = \{(0^{2^i}1^{2^i})^\infty, i \in \mathbb{N}\}$ whose first members are represented Figure 3. Any two members of $\mathcal{X}_\infty$ are at distance $\frac{1}{2}$. □

```
0101010101010101010101010101010101 ...
0011001100110011001100110011001100 ...
0000111100001111000011110000111100 ...
0000000011111111000000001111111100 ...
0000000000000000111111111111111100 ...
```

**Fig. 3.** first members of $\{(0^{2^i}1^{2^i})^\infty, i \in \mathbb{N}\}$

However, it remains open to find a set where configurations are all at distance strictly greater that one half, though for any $\varepsilon > 0$, one can find configurations whose relative distance is less than $\frac{1}{2} + \varepsilon$.

## References


1. Allouche, J.-P., Shallit, J.O.: Automatic Sequences - Theory, Applications, Generalizations. Cambridge University Press (2003)
2. Blanchard, F., Cervelle, J., Formenti, E.: Some results about the chaotic behavior of cellular automata. Theor. Comput. Sci. 349(3), 318–336 (2005)
3. Blanchard, F., Formenti, E., Kůrka, P.: Cellular automata in the Cantor, Besicovitch and Weyl Topological Spaces. Complex Systems 11, 107–123 (1999)
4. Cattaneo, G., Formenti, E., Margara, L., Mazoyer, J.: A Shift-invariant Metric on $S^{\mathbb{Z}}$ Inducing a Non-trivial Topology. In: Privara, I., Ružička, P. (eds.) MFCS 1997. LNCS, vol. 1295, pp. 179–188. Springer, Heidelberg (1997)
5. Durand, B.: Tilings and Quasiperiodicity. In: Degano, P., Gorrieri, R., Marchetti-Spaccamela, A. (eds.) ICALP 1997. LNCS, vol. 1256, pp. 65–75. Springer, Heidelberg (1997)
6. Formenti, E.: On the sensitivity of cellular automata in Besic ovitch spaces. Theoretical Computer Science 301(1-3), 341–354 (2003)
7. Hedlund, G.A.: Endomorphism and automorphism of the shift dynamical system. Mathematical System Theory 3, 320–375 (1969)
8. Semakov, N.V., Zinov'ev, V.A.: Equidistant q-ary codes with maximal distance and resolvable balanced incomplete block designs. Problemy Peredachi Informatsii 4(2), 3–10 (1968)