Cellular automata and substitutions in the edit-distance space

Firas Ben Ramdhane

Cellular automata and substitutions in the edit-distance space

Firas Ben Ramdhane[1]

[1]Aix Marseille Univ, CNRS, Centrale Marseille, I2M, Marseille, France
Sfax University, Faculty of Sciences of Sfax, AGTS, Sfax, Tunisia.
firas.ben-ramdhane@etu.univ-amu.fr.

Abstract: The Besicovitch and Weyl pseudo-distances are shift-invariant pseudo-metrics on the set of infinite sequences, that enjoy interesting properties and are suitable to study the dynamics of cellular automata. They correspond to the asymptotic behavior of the Hamming distance of longer and longer prefixes or factors. In this paper we replace Hamming distance by that of Levenshtein, with the aim of studying symbolic dynamical systems in their associated quotient space. We prove that every cellular automaton is Lipschitz with respect to this new distance, moreover, the shift-map is exactly the identity over those spaces. In addition, we show that, in the Besicovitch and Weyl spaces, substitutions are well-defined essentially only when they are uniform. However, we prove that in the new spaces associated to the Levenshtein distance, all substitutions are well-defined, and furthermore Lipschitz. Finally, we propose a general definitions of pseudo-metrics depending on the distance.

1 Introduction

In [BFK97] were studied the dynamics of cellular automata in the spaces of sequences endowed with the Besicovitch or Weyl pseudo-metrics, which are defined as asymptotics of the Hamming distance over prefixes or factors of the sequences. This corresponds to the $\bar{d}$-metric defined for ergodic purposes in [Orn74]. [Fel76], and independently [Kat77], proposed to replace the Hamming distance by the Levenshtein (or edit) distance from [Lev66], and get the $\bar{f}$-metric, which is useful in Kakutani equivalence theory. One can read some properties of the metric in [ORW82, Chapter 2], and a nice history of this notion in [KL17].

In this paper we will use some basic vocabulary of symbolic dynamical systems, for that we can cite [Kůr03] as a good reference for this theory.

We call an alphabet every finite set of symbols (or letters) it will be denoted A, a finite word over A is a finite sequence of letters in A, it is convenient to write a word as $u = u_0 \dots u_{|u|-1}$ to express u as the concatenation of the letters $u_0, u_1, \dots, u_{|u|-1}$ with $|u|$ representing the length of u, that is, the number of letters appearing in u. The unique word of length 0 is the empty word denoted by λ. An infinite word over an alphabet A is the concatenation of infinite letters and we write $x = x_0x_1x_2\dots$.

The set of all finite (resp. infinite) words over A is denoted by A^* (resp. $A^{\mathbb{N}}$) and for $n \in \mathbb{N}$, A^n is the set of words of length n over A. We use the notation $[\![i, j[\![$ for the set of integers $\{i, i+1, \dots, j-1\}$.

The shift map denoted σ is a function over $A^{\mathbb{N}}$, such that for $x \in A^{\mathbb{N}}$ we have $\sigma(x)_i = x_{i+1}$ for all $i \in \mathbb{N}$. A distance d over a set of finite words A^* is a function over $A^* \times A^*$ to $\mathbb{R}_+$ satisfying: separation, symmetry, and the triangle inequality. A pseudo-metric is an apllication satisfying the distance property except the separation property, distincts words can have zero distance. We are interested in distances defined between pairs of words of the same length. The prototypical example is the Hamming distance denoted by d_H, which counts the number of differences between two words, $d_H(u, v) = |\{\, i \in \mathbb{N} |\, u_i \neq v_i \}\,|$.

2 Besicovitch and Weyl spaces

In this section we recall the definition and topological properties of Besicovitch and Weyl spaces, then we introduce cellular automata in those spaces which are studied in several works. Finally we prove that not every substitution is well defined over those spaces.

2.1 Definitions and Topological properties

Definition 2.1 *The Besicovitch and Weyl pseudo-metrics, denoted here by $\mathfrak{C}_{d_H}$ and $\mathfrak{S}_{d_H}$ respectively, were defined in [BFK97] as follows:*

- $\mathfrak{C}_{d_H}(x, y) = \limsup_{l \to \infty} \dfrac{d_H(x_{[\![0,l[\![}, y_{[\![0,l[\![})}{l}, \ \ \forall x, y \in A^{\mathbb{N}}.$
- $\mathfrak{S}_{d_H}(x, y) = \limsup_{l \to \infty} \max_{k \in \mathbb{N}} \dfrac{d_H(x_{[\![k,k+l[\![}, y_{[\![k,k+l[\![})}{l}, \forall x, y \in A^{\mathbb{N}}.$

With a simple calculation, one can find that, for all $x, y \in A^{\mathbb{N}}$: $\mathfrak{C}_{d_H}(x, y) \leq \mathfrak{S}_{d_H}(x, y)$, and both of Besicovitch and Weyl pseudo-metrics are σ-invariant.

Since those two functions are pseudo-metrics and they are not distances, we factor the space of infinite words by the equivalence of zero distance, as mentioned in the following definitions:

Definitions 2.2 *Let A be an alphabet, and, let $X = A^{\mathbb{N}}$. The relation $x \sim_{\mathfrak{C}_{d_H}} y \Longleftrightarrow \mathfrak{C}_{d_H}(x, y) = 0$, (resp. $x \sim_{\mathfrak{S}_{d_H}} y \Longleftrightarrow \mathfrak{S}_{d_H}(x, y) = 0$) is an equivalence relation. The quotient space $X_{/\sim_{\mathfrak{C}_{d_H}}}$ (resp. $X_{/\sim_{\mathfrak{S}_{d_H}}}$) is a topological space called the **Besicovitch** (resp. **Weyl**) space denoted by $X_{\mathfrak{C}_{d_H}}$ (resp. $X_{\mathfrak{S}_{d_H}}$).*

We denote by $\tilde{x}$ the equivalence calss of $x \in X$ in the quotient space.

According to [BFK97], the Besicovitch space is pathwise-connected, infinite-dimension and complete topological space, but, it is neither separable nor locally compact. The Weyl space shares many properties with the Besicovitch space; one of the main differences is that it is not complete, according to [DI88].

2.2 Cellular automata

First of all, we recall that a cellular automaton of radius r is a map $F : A^{\mathbb{N}} \to A^{\mathbb{N}}$, such that there exists a map $f : A^r \to A$ such that for all $x \in A^{\mathbb{N}}$, $i \in \mathbb{N}$: $F(x)_i = f(x_{[\![i,i+r[\![})$. Necessary and sufficient conditions depending of the metric Cantor space, so that a function F is a cellular automaton, was given by Hedlund in [Hed69]. More precisely, a function $F : A^{\mathbb{N}} \to A^{\mathbb{N}}$ is a cellular automaton if and only if it is a continous function with respect to the metric Cantor space and shift-invariant (i.e. $F(\sigma(x)) = \sigma(F(x)),\ \forall x \in A^{\mathbb{N}}$). In [BFK97], it is shown that every cellular automaton induces a (well-defined) Lipshitz function over Besicovitch and Weyl spaces.

Definition 2.3 *A function $\tilde{F}$ is a cellular automaton on the Besicovitch (resp. Weyl) space if there exists a cellular automaton $G : A^{\mathbb{N}} \to A^{\mathbb{N}}$ such that $G \in \tilde{F}$, i.e.* $\forall x \in A^{\mathbb{N}}, \forall y \in \tilde{F}(\tilde{x}),\ \mathfrak{C}_{d_H}(G(x), y) = 0$ *(resp. $\mathfrak{S}_{d_H}(G(x), y) = 0$).*

[MS09, Theorem 13] gives a characterization similar to the Curtis-Hedlund-Lyndon of cellular automata by three conditions: shift invariance, a condition in terms of uniform continuity and a condition in terms of periodic configurations.

2.3 Substitution

In the first place, we recall some definitions used in the study of substitutions. For more details on substitutions we can refer [Fog02] and [BR10].

Definition 2.4 *Let A be an alphabet.*

1. *A* substitution *τ is a non-erasing morphism over A^*, i.e. τ replaces the letters of an alphabet A with non-empty finite words.*
2. *The function associated to a substitution τ is denoted $\overline{\tau}$ and defined on $A^{\mathbb{N}}$ by $\overline{\tau}(z) = \tau(z_0)\tau(z_1)\ldots, \forall z \in A^{\mathbb{N}}$.*

Definition 2.5 *A substitution τ is called* uniform *if for all $a, b \in A$, $|\tau(a)| = |\tau(b)|$. The length of a uniform substitution is $|\tau(a)|$ with $a \in A$.*

Now, we prove that the function associated to uniform substitutions induces a Lipschitz function on the Besicovitch (resp. Weyl) spaces.

Proposition 2.6 *For every uniform substitution τ with length $L \in \mathbb{N} \setminus \{0\}$, $\overline{\tau}$ is L-Lipschitz, in particular it is well-defined over X_B (resp. X_W)*

Proof: Let τ be a substitution and $L = \max_{a\in A} |\tau(a)|$. Let $x, y \in A^{\mathbb{N}}$ and $k \in \mathbb{N}$.
If $\overline{\tau}(z)_k \neq \overline{\tau}(y)_k$ then $\tau(z_{\lfloor k/L \rfloor}) \neq \tau(y_{\lfloor k/L \rfloor})$, hence $z_{\lfloor k/L \rfloor} \neq y_{\lfloor k/L \rfloor}$.
Then, for $l \in \mathbb{N}$ we have:

$$\begin{aligned} |\{k \in [\![0,l[\![\mid \overline{\tau}(z)_k \neq \overline{\tau}(y)_k\}| &\leq L \cdot \left|\left\{ k \in \left[\!\!\left[0, \left\lfloor \frac{l}{L} \right\rfloor \right[\!\!\right[\;\middle|\; z_k \neq y_k \right\}\right| \\ &\leq L \cdot |\{k \in [\![0,l[\![\mid z_k \neq y_k\}|. \end{aligned}$$

Therefore: $\mathfrak{C}_{d_H}(\overline{\tau}(z), \overline{\tau}(y)) \leq L \cdot \mathfrak{C}_{d_H}(z,y)$.
In conclusion, $\overline{\tau}$ is well-defined on the Besicovitch space and it is L-Lipschitz. ■

The following theorem shows that a non-constant function associated to a non-uniform substition could not be induced on the Besicovitch (resp. Weyl) space.

Theorem 2.7 *For every non-uniform substitution τ, $\overline{\tau}$ is well-defined on the Besicovitch space if and only if it is constant.*

Proof: If $\overline{\tau}$ is constant then $\mathfrak{C}_{d_H}(\overline{\tau}(x), \overline{\tau}(y)) = 0$ for all $x, y \in A^{\mathbb{N}}$. Hence $\overline{\tau}$ is well-defined on the Besicovitch space. Furthermore, the induced function is also constant on the quotient space.

We suppose now that $\overline{\tau}$ induces a function on the Besicovitch space, so, for all $x, y \in A^{\mathbb{N}}$, if $\mathfrak{C}_{d_H}(x,y) = 0$, then $\mathfrak{C}_{d_H}(\overline{\tau}(x), \overline{\tau}(y)) = 0$. We also suppose that τ is non-uniform, so there exists $a \neq b \in A$ such that $|\tau(a)| = |\tau(b)| + k$ with $k \in \mathbb{N} \setminus \{0\}$.
It is clear that, for any $z \in A^{\mathbb{N}}$ we have $\mathfrak{C}_{d_H}(az, bz) = 0$, hence:

$$\mathfrak{C}_{d_H}(\overline{\tau}(az), \overline{\tau}(bz)) = \mathfrak{C}_{d_H}(\overline{\tau}(z), \sigma^k(\overline{\tau}(z))) = 0.$$

Furthermore, for $z = a^\infty$ we have $\overline{\tau}(a^\infty)$ and $\sigma^k(\overline{\tau}(a^\infty))$ are periodic.

Since $\mathfrak{C}_{d_H}(\overline{\tau}(a^\infty), \sigma^k(\overline{\tau}(a^\infty))) = 0$ thus $\overline{\tau}(a^\infty) = \sigma^k(\overline{\tau}(a^\infty))$ due to the fact that $\mathfrak{C}_{d_H}(x,y) > 0$ for every periodic $x, y \in A^{\mathbb{N}}$ according to [CFMM97, Proposition 3]. Therefore $\overline{\tau}(a^\infty)$ is $|\tau(a)|$-periodic and k-periodic configuration. Hence, according to Fine and Wilf's theorem [FW65], there exists $v \in A^*$ and $j \in \mathbb{N} \setminus \{0\}$ such that $|v|$ divides k and $\tau(a) = v^j$. Then, for all $d \in A \setminus \{a\}$ we take $x^d \in A^{\mathbb{N}}$ periodic starting with ad and we find $\tau(d) = v^m$ with $m \in \mathbb{N} \setminus \{0\}$. We then deduce that there exists $v \in A^*$, for any $c \in A$ there exists $m \in \mathbb{N} \setminus \{0\}$ such that $\tau(c) = v^m$.
In conclusion, $\overline{\tau}$ is constant. ■

Examples 2.8 *Let τ be the substitution defined on $A = \{0,1\}$ by $\tau(0) = 01$ and $\tau(1) = 1$. Then we get $x = 0^\infty$ and $y = 10^\infty$.*
We can remark that, $\mathfrak{C}_{d_H}(x,y) = 0$. We have also, $\overline{\tau}(x) = (01)^\infty$ and $\overline{\tau}(y) = (10)^\infty$. Then $\mathfrak{C}_{d_H}(\tau(x), \tau(y)) = \frac{1}{2}$. Hence, $\widetilde{\overline{\tau}(x)} \neq \widetilde{\overline{\tau}(y)}$, despite that $\tilde{x} = \tilde{y}$.

It is well known now that the Besicovitch (resp. Weyl) pseudo-metric and cellular automata and uniform substitutions fit very well together. But, we find that, this is not true in the case of non-uniform substitutions. For that, we introduce in the next section a new shift-invariant pseudo-metric, and we find that both of cellular automata and all substitutions are well defined on the metric topology induced by this pseudo-metric.

3 Levenshtein distance

Another distance used in symbolic dynamical systems is the Levenshtein distance [Lev66], it may also be referred to as the edit distance. It used extensively for information theory, linguistics, word algorithmics, ... This distance can be considered as a variant of the Hamming distance. We show in particular that the Hamming distance is an upper bound for the Levenshtein distance. It depends on the minimum number of edit operations required to change one word into the other. The edit operations are defined over A^* as follows: for $u \in A^*$, $a \in A$ and $i \in [\![0, |u|[\![$:

- The deletion of a letter: $D_i(u) = u_0u_1 \dots u_{i-1}u_{i+1} \dots u_{|u|-1}$
- The substitution of a letter by another one: $S_i^a(u) = u_0u_1 \dots u_{i-1}au_{i+1} \dots u_{|u|-1}$
- The insertion of a letter: $I_i^a(u) = u_0 \dots u_{i-1}au_iu_{i+1} \dots u_{|u|-1}$.

Using these notations, We can write the Levenshtein distance as follows:

Definition 3.1 *The Levenshtein distance between words u and v, denoted by $d_L(u,v)$, is defined as the minimal $n \in \mathbb{N}$ such that $T_1 \circ \dots \circ T_l(u) = v$ for some $T_k \in \{D_i | i\} \sqcup \{S_i^a | a \in A, i\} \sqcup \{I_i^a | a \in A, i\}$, for $k \in [\![1, l]\!]$.*
There are constraints on the sites i where each operation is performed, depending on the current length of the image of word u, but for readability, we do not write them explicitly.

The definition is not changed in its symmetric version, that is, allowing to modify both u and v to get to an equality, as it is sometimes seen in the literature.

Examples 3.2 *Let $A = \{0, 1\}$.*

1. *For $u = 010101$ and $v = 101010$, we have : $d_L(u,v) = 2$. In fact, $D_0(u) = 10101$ (we delete the letter of index 0 in u), then u became equals to 10101, now we add the letter 0 in the end of the word u ($I_5^0(u) = 101010$) and we find $I_5^0 \circ D_0(u) = 101010 = v$.*
2. *For $u = 0000$ and $v = 0001$, we have : $d_L(u,v) = 1$. In fact, it is enough to change $u_3 = 0$ by 1 i.e. $S_3^1(u) = 0001 = v$.*

Following the idea behind the Besicovitch and Weyl pseudo-metrics, we define two pseudo-metrics associated to the Levenshtein distance as follows:

Definitions 3.3 *Let A be an alphabet.*

1. *The* centred pseudo-metric *associated to the Levenshtein distance is:*

$$\mathfrak{C}_{d_L}(x,y) = \limsup_{l \to \infty} \frac{d_L(x_{[\![0,l[\![}, y_{[\![0,l[\![})}{l}, \quad \forall x, y \in A^{\mathbb{N}}.$$

2. *The* sliding pseudo-metric *associated to the Levenshtein distance is:*

$$\mathfrak{S}_{d_L}(x,y) = \limsup_{l\to\infty} \max_{k\in\mathbb{N}} \frac{d_L(x_{[\![k,k+l[\![}, y_{[\![k,k+l[\![})}{l}, \quad \forall x,y \in A^{\mathbb{N}}.$$

Since those two functions are pseudo-metrics and they are not distances, we quotient the space of infinite words by the equivalence of zero distance to find two metric spaces called centred (resp. sliding) space associated to Levenshtein distance denoted by $X_{\mathfrak{C}_{d_L}}$ (resp. $X_{\mathfrak{S}_{d_L}}$) where $X = A^{\mathbb{N}}$.

Proposition 3.4 *The Levenshtein distance satisfies the following properties :*

1. $d_L(u,v) \leq d_H(u,v)$, $\forall u,v \in A^*$.
2. $d_L(uu',vv') \leq d_L(u,v) + d_L(u',v')$, $\forall u,v \in A^n$, $\forall u',v' \in A^m$, *with* $n,m \in \mathbb{N}$. *This property is called* subadditivity.
3. $\mathfrak{C}_{d_L}(x,y) \leq \mathfrak{C}_{d_H}(u,v)$, $\forall x,y \in A^{\mathbb{N}}$ *and* $\mathfrak{S}_{d_L}(x,y) \leq \mathfrak{S}_{d_H}(u,v)$, $\forall x,y \in A^{\mathbb{N}}$

Proof:

1. Let $d_H(u,v) = n$, then there exists n difference between u and v. If we change the elements where u and v are different and let them equals, we used then n substitution to get the equality between u and v. So $d_L(u,v) \leq n = d_H(u,v)$.
2. Let $u,v \in A^n$, $u',v' \in A^m$. Denote $d_L(u,v) = p$ and $d_L(u',v') = q$ with $p,q \in \mathbb{N}\setminus\{0\}$, then the minimum number of operations used to let u and v equals is p and the minimum number of operations used to let u' and v' equals is q. Hence the minimum number of operations used to let uu' and vv' equals is less or equal to $p+q$. Therefore : $d_L(uu',vv') \leq p+q = d_L(u,v) + d_L(u',v')$.
3. Both inequalities are deduced from $d_L(u,v) \leq n = d_H(u,v)$.

■

Proposition 3.5 *The shift map over the centred (resp. sliding) space associated to the Levenshtein distance is exactly the identity map. In particular the pseudo-metrics associated to the Levenshtein distance are σ-invariant.*

Proof: Let $\tilde{x} \in X_{\mathfrak{C}_{d_L}}$. If $x \in \tilde{x}$, then: $d_L(x_{[\![0,l[\![}, \sigma(x)_{[\![0,l[\![}) = d_L(x_{[\![0,l[\![}, x_{[\![1,1+l[\![}) \leq 2$.
We just delete the first letter of $x_{[\![0,l[\![}$ (we find $x_{[\![1,l[\![}$) and then we insert x_{l+1} to find the equality. Hence :

$$\mathfrak{C}_{d_L}(x,\sigma(x)) = \limsup_{l\to\infty} \frac{d_L(x_{[\![0,l[\![}, x_{[\![1,1+l[\![})}{l} \leq \limsup_{l\to\infty} \frac{2}{l} = 0.$$

So $x \in \widetilde{\sigma(x)} = \tilde{\sigma}(\tilde{x})$. And then, $\tilde{x} \subseteq \tilde{\sigma}(\tilde{x})$.
If $z \in \tilde{\sigma}(\tilde{x})$, then $\mathfrak{C}_{d_L}(z,x) \leq \mathfrak{C}_{d_L}(z,\sigma(x)) + \mathfrak{C}_{d_L}(\sigma(x),x) = 0$. Hence $\mathfrak{C}_{d_L}(z,x) = 0$.
So $z \in \tilde{x}$, therefore $\tilde{\sigma}(\tilde{x}) \subseteq \tilde{x}$. In conclusion, for all $x \in A^{\mathbb{N}}$ we have, $\tilde{\sigma}(\tilde{x}) = \tilde{x}$. ■

Since every class is invariant by shift, dynamical systems over this space can be considered as acting on shift orbits.

3.1 Cellular automata

Now, we aim to prove that every cellular automaton is well defined over this new quotient space and furthermore Lipschitz.

Definition 3.6 *Let $F: A^{\mathbb{N}} \to A^{\mathbb{N}}$ be a cellular automaton with radius r and a local function $f: A^r \to A$. We denote by f^* the function defined over A^* to itself by:*

$$f^*(u) = f(u_{[\![0,r[\![})f(u_{[\![1,1+r[\![}) \dots f(u_{[\![l,l+r[\![}),$$

for u such that $|u| \geq r$ and $f^*(u) = \lambda$ if $|u| < r$.

Lemma 3.7 *For all $u, v \in A^*$ and $w, w' \in A^*$ we have: $d_L(uwv, uw'v) \leq max(|w|, |w'|)$.*

Proof: Let us suppose that $|w| \leq |w'|$. We can remark that we can do $|w|$ substitution for the word uwv to get $uw'_{[\![0,|w|[\![}v$ then we do $|w'| - |w|$ insertion of the elements of $w'_{[\![|w|,|w'|[\![}$ to get $uw'v$. So we find $|w'|$ operations, to make uwv equals to $uw'v$. Hence : $d_L(uwv, uw'v) \leq |w'| = max(|w|, |w'|)$. ■

Lemma 3.8 *Let F be a cellular automaton with radius r and local function f. Then for all $u \in A^{r+1}A^*, v \in A^rA^*$, and $T \in \{D_i | i\} \sqcup \{S_i^a | a \in A, i\} \sqcup \{I_i^a | a \in A, i\}$.*

Proof: Let $u \in A^{r+1}A^*, v \in A^rA^*$, and let $n = d_L(f^*(u), f^*(v))$. By the triangular inequality, we have : $d_L(f^*(T_i(u)), f^*(v)) \leq d_L(f^*(T_i(u)), f^*(u)) + d_L(f^*(u), f^*(v))$. So it is enough to show that $d_L(f^*(T_i(u)), f^*(u)) \leq r+1$. Let $f^*(u) = u'wv$, where $u' = f^*(u_{[\![0,i[\![})$, $w = f^*(u_{[\![p,i+r[\![})$ and $v = f^*(u_{[\![i+1,|u|-1[\![})$, with $p = \min\{i-r, 0\}$. One can note that :

- $f^*(S_i^a(u)) = u'w'v$ where $w' = f^*(u_{[\![p,i[\![}au_{[\![i+1,i+r[\![})$
- $f^*(I_i^a(u)) = u'w'v$ where $w' = f^*(u_{[\![p,i[\![}au_{[\![i,i+r[\![})$
- $f^*(D_i(u)) = u'w'v$ where $w' = f^*(u_{[\![p,i[\![}u_{[\![i+1,i+r[\![})$

In these three cases, we are in the case of the previous lemma, with $|w| = r$ and $r - 1 \leq |w'| \leq r+1$, which gives the wanted inequality. ■

Lemma 3.9 *Let F be a cellular automaton with radius r and local function f. Then for all $u, v \in A^rA^*$ such that $d_L(u, v) \leq |u| - r$, we have:*

$$d_L(f^*(u), f^*(v)) \leq (r+1)d_L(u, v).$$

Proof: Let $u, v \in A^rA^*$ such that $n = d_L(u, v) \leq |u| - r$. By definition, we can write $u = T_1 \circ \dots \circ T_n(v)$ for some operations $T_k \in \{D_i | i\} \sqcup \{S_i^a | a \in A, i\} \sqcup \{I_i^a | a \in A, i\}$. Since the length decreases by at most 1, note that the assumption implies that no word in the sequence of operations has a length smaller than r. We can prove by induction on k decreasing from n to 0 that:

$$d_L(f^*(T_{k+1} \circ \dots \circ T_{i_n}(v)), f^*(v)) \leq (r+1)(n-k).$$

The case $k = n$ is trivial. Now suppose that $d_L(f^*(T_{k+1} \circ \ldots \circ T_{i_n}(v)), f^*(v)) \leq (r+1)(n-k)$ for some k. By Lemma 3.8, $d_L(f^*(T_k \circ \ldots \circ T_{i_n}(v)), f^*(v)) \leq r+1+(r+1)(n-k) = (r+1)(n-k+1)$, which is exactly the next step if the induction hypothesis. We obtain (for $k = 0$) the claimed statement. ■

Theorem 3.10 *Every cellular automaton F with radius r is $(r+1)$-Lipschitz with respect to $\mathfrak{C}_{d_L}$. In particular it is well defined on the quotient space.*

Proof: Let $x, y \in A^{\mathbb{N}}$.

<u>*Case 1.*</u> If $\mathfrak{C}_{d_L}(x,y) = 1$, then $\mathfrak{C}_{d_L}(F(x), F(y) \leq \mathfrak{C}_{d_L}(x,y) = 1$.

<u>*Case 2.*</u> If $\mathfrak{C}_{d_L}(x,y) < 1$. Then : $\limsup_{l\to\infty} \dfrac{d_L(x_{[\![0,l[\![}, y_{[\![0,l[\![})}{l} < 1$.

Hence, there exists $N > 0$ such that for all $l > N$, we have: $\dfrac{d_L(x_{[\![0,l[\![}, y_{[\![0,l[\![})}{l} < 1 - \frac{r}{l}$.
Therefore, we find that $d_L(x_{[\![0,l[\![}, y_{[\![0,l[\![}) < l - r$. Applying Lemma 3.9, we deduce:

$$d_L(f^*(x_{[\![0,l[\![}), f^*(y_{[\![0,l[\![})) \leq (r+1) d_L(x_{[\![0,l[\![}, y_{[\![0,l[\![}).$$

Deviding by l and getting the upper limit, we find our result. ■

3.2 Substitutions

Finally, we prove that, contrariwise the Besicovitch and Weyl spaces, all substitutions are well-defined on the quotient space with respect to the pseudo-metric associated to the Levenshtein distance, as mentioned in the following theorem:

Theorem 3.11 *For every substitution τ, $\overline{\tau}$ is well-defined over $X_{\mathfrak{C}_{d_L}}$.
Furthermore it is L-Lipschitz with $L = \max_{a\in A} |\tau(a)|$.*

Proof: Let τ be a substitution and $L = \max_{a\in A} |\tau(a)|$. Let $x, y \in A^{\mathbb{N}}$ and $l \in \mathbb{N}$. We assume that $d_L(x_{[\![0,l[\![}, y_{[\![0,l[\![}) = n$. Then there exists a sequence of operations $T_k \in \{D_i | i\} \sqcup \{S_i^a | a \in A, i\} \sqcup \{I_i^a | a \in A, i\}$ for $1 \leq k \leq n$ such that:
$T_1 \circ \ldots \circ T_n(x_{[\![0,l[\![}) = y_{[\![0,l[\![}$. Hence : $\tau(T_1 \circ \ldots \circ T_n(x_{[\![0,l[\![})) = \tau(y_{[\![0,l[\![})$.
By an argument similar to that of Lemma 3.9, one can show that for every T, there exist $T'_1, ..., T'_{n'}$, with $n' \leq L$, such that $\tau(T(u)) = T'_1 \circ ... \circ T'_{n'} \tau(u)$. Then, a direct induction gives that there exist n' operations $(T'_k)_{1\leq k\leq n'}$ with $n' \leq nL$ such that: $T'_1 \circ \ldots \circ T'_{n'}(\tau(x_{[\![0,l[\![})) = \tau(y_{[\![0,l[\![})$. Therefore :

$$d_L(\tau(x_{[\![0,l[\![}), \tau(y_{[\![0,l[\![})) \leq n' \leq Ln = L \cdot d_L(x_{[\![0,l[\![}, y_{[\![0,l[\![}).$$

And since: $d_L(\tau(x)_{[\![0,l[\![}, \tau(y)_{[\![0,l[\![}) \leq d_L(\tau(x_{[\![0,l[\![}), \tau(y_{[\![0,l[\![}))$, then:

$$d_L(\tau(x)_{[\![0,l[\![}, \tau(y)_{[\![0,l[\![}) \leq L \cdot d_L(x_{[\![0,l[\![}, y_{[\![0,l[\![}).$$

Dividing by l and passing to the upper limit, we find: $\mathfrak{C}_{d_L}(\overline{\tau}(x), \overline{\tau}(y)) \leq L \cdot \mathfrak{C}_{d_L}(x,y)$.
In conclusion, $\overline{\tau}$ is L-Lipschitz with respect to $\mathfrak{C}_{d_L}$. ■

4 Conclusion

In this paper, we construct a metric which induces a non trivial topology and makes the shift map equal to the identity over this space. This topology turns out to be a suitable playground for the study of the dynamical behavior of substitutions and cellular automata, since both of them are Lipschitz functions. This construction was made only by changing the Hamming distance with the Levenshtein distance. We suggest global definitions for those pseudo-metrics, where we can change the Hamming distance by any other distance defined on the set of finite words, as follows :

Definitions 4.1 *For a distance d over $A^* \times A^*$, we define the centred pseudo-metric denoted by $\mathfrak{C}_d$ and the sliding pseudo-metric denoted by $\mathfrak{S}_d$ as follows:*

- $\mathfrak{C}_d(x,y) = \limsup_{l\to\infty} \dfrac{d(x_{[\![0,l[\![}, y_{[\![0,l[\![})}{max_{u,v\in A^l} d(u,v)}, \ \forall x,y \in A^{\mathbb{N}}.$

- $\mathfrak{S}_d(x,y) = \limsup_{l\to\infty} \max_{k\in\mathbb{N}} \dfrac{d(x_{[\![k,k+l[\![}, y_{[\![k,k+l[\![})}{max_{u,v\in A^l} d(u,v)}, \ \forall x,y \in A^{\mathbb{N}}.$

A relevant question is now the following: Which properties of distance d make CA or substitutions well-defined in the corresponding pseudometrics?
Generalizations exist of Besicovitch pseudometrics over groups (see for instance [LS16, CGN20]). An interesting work would be to generalize more of these metrics to this setting.

References

[BFK97] François Blanchard, Enrico Formenti, and Petr Kůrka. Cellular automata in the Cantor, Besicovitch, and Weyl topological spaces. *Complex Systems*, 11:107–123, 1997. 00000.

[BR10] Valérie Berthé and Michel Rigo, editors. *Combinatorics, Automata, and Number Theory*, volume 12. Cambridge University Press, 2010.

[CFMM97] Gianpiero Cattaneo, Enrico Formenti, Luciano Margara, and Jacques Mazoyer. A shift-invariant metric on s zz inducing a non-trivial topology. In *International Symposium on Mathematical Foundations of Computer Science*, pages 179–188. Springer, 1997.

[CGN20] Silvio Capobianco, Pierre Guillon, and Camille Noûs. A Characterization of Amenable Groups by Besicovitch Pseudodistances. In Hector Zenil, editor, *Cellular Automata and Discrete Complex Systems*, pages 99–110, Cham, 2020. Springer International Publishing. 00000.

[DI88] Tomasz Downarowicz and Anzelm Iwanik. Quasi-uniform convergence in compact dynamical systems. *Studia Mathematica*, 89(1):11–25, 1988.

[Fel76] J. Feldman. New K-automorphisms and a problem of Kakutani. *Israel Journal of Mathematics*, 24(1):16–38, March 1976. 00000.

[Fog02] N Pytheas Fogg. *Substitutions in dynamics, arithmetics and combinatorics*. Springer Science & Business Media, 2002.

[FW65] Nathan J Fine and Herbert S Wilf. Uniqueness theorems for periodic functions. *Proceedings of the American Mathematical Society*, 16(1):109–114, 1965.

[Hed69] Gustav A Hedlund. Endomorphisms and automorphisms of the shift dynamical system. *Mathematical systems theory*, 3(4):320–375, 1969.

[Kat77] A B Katok. Monotone Equivalence in Ergodic Theory. *Mathematics of the USSR-Izvestiya*, 11(1):99–146, February 1977. 00072.

[KL17] Dominik Kwietniak and Martha Łącka. Feldman-Katok pseudometric and the GIKN construction of nonhyperbolic ergodic measures. *arXiv:1702.01962 [math]*, February 2017. 00000 arXiv: 1702.01962.

[Kůr03] Petr Kůrka. *Topological and symbolic dynamics*. Société mathématique de France Paris, 2003.

[Lev66] Vladimir I Levenshtein. Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady*, volume 10, pages 707–710, 1966.

[LS16] Martha Łącka and Marta Straszak. Quasi-uniform Convergence in Dynamical Systems Generated by an Amenable Group Action. *arXiv:1610.09675 [math]*, October 2016. 00001 arXiv: 1610.09675.

[MS09] Johannes Müller and Christoph Spandl. A Curtis–Hedlund–Lyndon theorem for Besicovitch and Weyl spaces. *Theoretical Computer Science*, 410(38-40):3606–3615, September 2009. 00005.

[Orn74] Donald S. Ornstein. *Ergodic theory, Randomness, and Dynamical Systems*. Number 5 in Yale Mathematical Monographs. Yale University Press, 1974. 00000.

[ORW82] Donald S. Ornstein, Daniel J. Rudolph, and Benjamin Weiss. *Equivalence of Measure Preserving Transformations*. Number 262 in Memoirs of the AMS. AMS, May 1982. 00165.