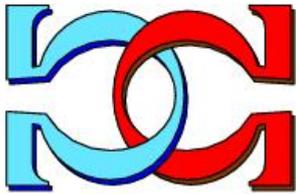
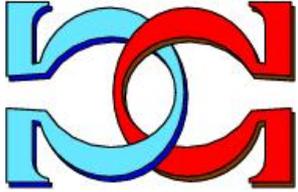
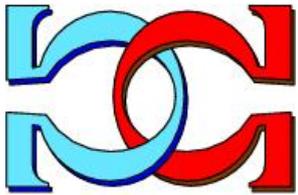


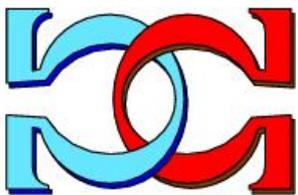
**CDMTCS  
Research  
Report  
Series**



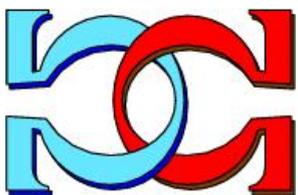
**Subword Metrics for Infinite  
Words**



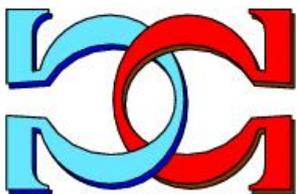
**Stefan Hoffmann**



**Ludwig Staiger**  
Martin-Luther-Universität  
Halle-Wittenberg



CDMTCS-484  
May 2015



Centre for Discrete Mathematics and  
Theoretical Computer Science

# Subword Metrics for Infinite Words

*Stefan Hoffmann*<sup>\*</sup>

Institut für Mathematik

and

*Ludwig Staiger*<sup>†</sup>

Institut für Informatik

Martin-Luther-Universität Halle-Wittenberg

D–06099 Halle (Saale), Germany

## Abstract

The space of one-sided infinite words plays a crucial rôle in several parts of Theoretical Computer Science. Usually, it is convenient to regard this space as a metric space, the Cantor-space. It turned out that for several purposes topologies other than the one of the Cantor-space are useful, e.g. for studying fragments of first-order logic over infinite words or for a topological characterisation of random infinite words.

Continuing the work of [SS10], here we consider two different refinements of the Cantor-space, given by measuring common factors, and common factors occurring infinitely often. In particular we investigate the relation of these topologies to the sets of infinite words definable by finite automata, that is, to regular  $\omega$ -languages

*Keywords:* metric spaces;  $\omega$ -words; subwords; shift-invariance; subword complexity

---

<sup>\*</sup>email: [st.hoffmann@student.uni-halle.de](mailto:st.hoffmann@student.uni-halle.de)

<sup>†</sup>email: [ludwig.staiger@informatik.uni-halle.de](mailto:ludwig.staiger@informatik.uni-halle.de)

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Notation and Preliminaries</b>	<b>2</b>
<b>3</b>	<b>The CANTOR Topology and Regular <math>\omega</math>-languages</b>	<b>3</b>
3.1	Basic properties of the CANTOR topology . . . . .	3
3.2	Regular $\omega$ -languages . . . . .	5
<b>4</b>	<b>Topologies Defined by Subword Metrics</b>	<b>6</b>
4.1	Shift-invariance . . . . .	6
4.2	Balls in $(X^\omega, \rho_I)$ and $(X^\omega, \rho_\infty)$ . . . . .	7
4.3	Non-preservation of regular $\omega$ -languages . . . . .	9
<b>5</b>	<b>Completeness and Compactness</b>	<b>10</b>
<b>6</b>	<b>Subword Complexity</b>	<b>11</b>

## 1 Introduction

The space of one-sided infinite words plays a crucial rôle in several parts of Theoretical Computer Science (see the surveys [St97, Th90]). Usually, it is convenient to regard this space as a topological space provided with the CANTOR topology. This topology can be also considered as the natural continuation of the left topology of the prefix relation on the space of finite words (cf. [CJ09]).

It turned out that for several purposes other topologies on the space of infinite words are also useful [Re86, St87], e.g. for investigations in first-order logic [DK09], to characterise the set of random infinite words [CM03] or the set of disjunctive infinite words [St05] and to describe the converging behaviour of not necessarily hyperbolic iterative function systems [FS01, St03].

Most of these approaches use topologies on the space of infinite words which are refinements of the CANTOR topology showing a certain kind of shift invariance. In [SS10] a unified treatment of those shift invariant topologies is given, and here we built on this work, introducing two new topologies arising naturally from the consideration of finite subwords occurring in infinite words.

## 2 Notation and Preliminaries

We introduce the notation used throughout the paper. By  $\mathbb{N} = \{0, 1, 2, \dots\}$  we denote the set of natural numbers. Let  $X$  be a finite alphabet of cardinality

$|X| \geq 2$ , and  $X^*$  be the set (monoid) of words on  $X$ , including the *empty word*  $e$ , and  $X^\omega$  be the set of infinite sequences ( $\omega$ -words) over  $X$ . For  $w \in X^*$  and  $\eta \in X^* \cup X^\omega$  let  $w \cdot \eta$  be their *concatenation*. This concatenation product extends in an obvious way to subsets  $W \subseteq X^*$  and  $P \subseteq X^* \cup X^\omega$ . For a language  $W$  let  $W^* := \bigcup_{i \in \mathbb{N}} W^i$  be the *submonoid* of  $X^*$  generated by  $W$ , and by  $W^\omega := \{w_1 \cdots w_i \cdots : w_i \in W \setminus \{e\}\}$  we denote the set of infinite strings formed by concatenating words in  $W$ . Furthermore  $|w|$  is the *length* of the word  $w \in X^*$  and  $\mathbf{pref}(P)$  ( $\mathbf{infix}(P)$ ) is the set of all finite prefixes (infixes) of strings in  $P \subseteq X^* \cup X^\omega$ . We shall abbreviate  $w \in \mathbf{pref}(\eta)$  ( $\eta \in X^* \cup X^\omega$ ) by  $w \sqsubseteq \eta$ . If  $\xi \in X^\omega$  by  $\mathbf{infix}^\infty(\xi) \subseteq \mathbf{infix}(\xi)$  we denote the set of infixes occurring infinitely often in  $\xi$ .

Further we denote by  $P/w := \{\eta : w \cdot \eta \in P\}$  the *left derivative* or *state* of the set  $P \subseteq X^* \cup X^\omega$  generated by the word  $w$ . We refer to  $P$  as *finite-state* provided the set of states  $\{P/w : w \in X^*\}$  is finite. It is well-known that a language  $W \subseteq X^*$  is finite state if and only if it is accepted by a finite automaton, that is, it is a regular language.<sup>1</sup>

In the case of  $\omega$ -languages *regular  $\omega$ -languages*, that is,  $\omega$ -languages accepted by finite automata, are the finite unions of sets of the form  $W \cdot V^\omega$ , where  $W$  and  $V$  are regular languages (cf. e.g. [St97]). Every regular  $\omega$ -language is finite-state, but, as it was observed in [Tr62], not every finite-state  $\omega$ -language is regular (cf. also [St83]).

It is well-known that the families of regular or finite-state  $\omega$ -languages are closed under Boolean operations (see [PP04, St97, Th90, Th97] or [St83]).

### 3 The CANTOR Topology and Regular $\omega$ -languages

In this section we list some properties of the CANTOR topology on  $X^\omega$  and regular  $\omega$ -languages (see [St97, Th90]).

#### 3.1 Basic properties of the CANTOR topology

We consider the space of infinite words ( $\omega$ -words)  $X^\omega$  as a metric space with metric  $\rho$  defined as follows

$$\rho(\xi, \eta) := \begin{cases} 0, & \text{if } \xi = \eta, \text{ and} \\ \sup\{r^{1-|w|} : w \in \mathbf{pref}(\xi) \Delta \mathbf{pref}(\eta)\} & \text{if } \xi \neq \eta. \end{cases} \quad (1)$$

<sup>1</sup>Observe that the relation  $\sim_P$  defined by  $w \sim_P v$  iff  $P/w = P/v$  is the NERODE right congruence of  $P$ .

Here  $r > 1$  is a real number<sup>2</sup>,  $\Delta$  denotes the symmetric difference of sets and we set  $\sup \emptyset := 0$ , that is,  $\rho(\xi, \eta) = 0$  if and only if  $\xi = \eta$ .

Since  $\mathbf{pref}(\xi) \Delta \mathbf{pref}(\eta) \subseteq (\mathbf{pref}(\xi) \Delta \mathbf{pref}(\zeta)) \cup (\mathbf{pref}(\zeta) \Delta \mathbf{pref}(\eta))$ , the metric  $\rho$  satisfies the ultra-metric inequality

$$\rho(\xi, \eta) \leq \max\{\rho(\xi, \zeta), \rho(\zeta, \eta)\}.$$

A subset  $E \subseteq X^\omega$  is *open* if for every  $\xi \in E$  there is an  $\varepsilon > 0$  such that  $\eta \in E$  for all  $\eta$  with  $\rho(\xi, \eta) < \varepsilon$ . Complements of open sets are called *closed*. The smallest closed set containing a given set  $F \subseteq X^\omega$ ,  $\mathcal{C}(F)$ , is referred to as the *closure of F*.

$\mathbf{G}_\delta$ -sets are countable intersections of open sets and  $\mathbf{F}_\sigma$ -sets are countable unions of closed sets. In a metric space every open set is an  $\mathbf{F}_\sigma$ -set, and every closed set is a  $\mathbf{G}_\delta$ -set.

We list some further well-known properties of the metric space  $(X^\omega, \rho)$ .

**Property 1** *The following is true.*

1. *The non-empty sets  $w \cdot X^\omega$  are open balls with radius  $r^{-|w|}$  in the metric space  $(X^\omega, \rho)$ .<sup>3</sup> These balls are simultaneously closed.*
2. *Open sets in  $(X^\omega, \rho)$  are of the form  $W \cdot X^\omega$  where  $W \subseteq X^*$ .*
3. *A subset  $E \subseteq X^\omega$  is open and closed (clopen) in  $(X^\omega, \rho)$  if and only if  $E = W \cdot X^\omega$  where  $W \subseteq X^*$  is finite.*
4. *A subset  $F \subseteq X^\omega$  is closed in  $(X^\omega, \rho)$  if and only if  $F = \{\xi : \mathbf{pref}(\xi) \subseteq \mathbf{pref}(F)\}$ .*
5. *The closure of  $F$  satisfies*

$$\begin{aligned} \mathcal{C}(F) &:= \{\xi : \xi \in X^\omega \wedge \mathbf{pref}(\xi) \subseteq \mathbf{pref}(F)\} \\ &= \bigcap_{n \in \mathbb{N}} (\mathbf{pref}(F) \cap X^n) \cdot X^\omega. \end{aligned}$$

The space  $(X^\omega, \rho)$  is a *complete* space, that is, every sequence<sup>4</sup>  $(\xi_i)_{i \in \mathbb{N}}$  where  $\rho(\xi_j, \xi_k) < r^{-i}$  whenever  $i \leq j, k$  converges to some  $\xi \in X^\omega$ . Moreover,  $(X^\omega, \rho)$  is a *compact* space, that is, for every family of open sets  $(E_i)_{i \in J}$  such that  $\bigcup_{i \in J} E_i = X^\omega$  there is a finite sub-family  $(E_i)_{i \in J'}$  satisfying  $\bigcup_{i \in J'} E_i = X^\omega$ .

<sup>2</sup>It is convenient to choose  $r = |X|$ . Then every ball of radius  $r^{-n}$  is partitioned into exactly  $r$  balls of radius  $r^{-(n+1)}$

<sup>3</sup>Observe that  $e \notin \mathbf{pref}(\xi) \Delta \mathbf{pref}(\eta)$  and Eq. (1) imply  $\rho(\xi, \eta) = \inf\{r^{-|w|} : w \sqsubset \xi \wedge w \sqsubset \eta\}$ .

<sup>4</sup>Those sequences are usually referred to as CAUCHY sequences.

### 3.2 Regular $\omega$ -languages

As a last part of this section we mention some facts on regular  $\omega$ -languages known from the literature, e.g. [PP04, St97, Th90]. Regular  $\omega$ -languages are well-known for being the  $\omega$ -languages definable by finite automata. We will not refer to this feature, instead we list some basic properties of this family of  $\omega$ -languages.

The first one shows among other properties the importance of ultimately periodic  $\omega$ -words. Denote by  $\text{Ult} := \{w \cdot v^\omega : w, v \in X^* \setminus \{e\}\}$  the set of ultimately periodic  $\omega$ -words.

**Theorem 1 (Büchi [Bü60])** *The family of regular  $\omega$ -languages is a Boolean algebra, and if  $F \subseteq X^\omega$  is regular, then  $u \cdot F$  and  $F / w$  are also regular.*

*Every non-empty regular  $\omega$ -language contains an ultimately periodic  $\omega$ -word, and regular  $\omega$ -languages  $E, F \subseteq X^\omega$  coincide if and only if  $E \cap \text{Ult} = F \cap \text{Ult}$ .*

For regular  $\omega$ -languages we have the following topological characterisations analogous to Property 1.

**Property 2** *Let  $F \subseteq X^\omega$  be regular and  $E \subseteq X^\omega$  be finite-state. Then in CAN-TOR topology the following hold true.*

1.  $F$  is open if and only if  $F = W \cdot X^\omega$  where  $W \subseteq X^*$  is a regular language.
2.  $F \subseteq X^\omega$  is closed if and only if  $F = \{\xi : \text{pref}(\xi) \subseteq \text{pref}(F)\}$  and  $\text{pref}(F)$  is regular.
3.  $\text{pref}(E)$  is a regular language.
4.  $\mathcal{C}(E)$  is a regular  $\omega$ -language.

Finally, we provide an example of a regular  $\omega$ -language which is not a  $\mathbf{G}_\delta$ -set and a necessary and sufficient topological condition when finite-state  $\omega$ -languages are regular.

**Example 1 (Landweber [La69])** For  $u \in X^* \setminus \{e\}$  the  $\omega$ -language  $X^* \cdot u^\omega$  is regular, an  $\mathbf{F}_\sigma$ -set but not a  $\mathbf{G}_\delta$ -set.  $\square$

**Theorem 2 ([St83])** *Every finite-state  $\omega$ -language in the class  $\mathbf{F}_\sigma \cap \mathbf{G}_\delta$  is a Boolean combination of regular  $\omega$ -languages open in  $(X^\omega, \rho)$ , thus, in particular, a regular  $\omega$ -language.*

## 4 Topologies Defined by Subword Metrics

It was shown that regular  $\omega$ -languages are closely related to the (asymptotic) subword complexity of infinite words (cf. [St93, Section 5] and [St12]). Therefore, as other refinements of the CANTOR topology we introduce two topologies defined via metrics on  $X^\omega$  which are based on the sets of subwords occurring or occurring infinitely often in the  $\omega$ -words, respectively.

### Definition 1 (Subword metrics)

$$\begin{aligned}\rho_I(\xi, \eta) &:= \sup\{r^{1-|w|} : w \in (\mathbf{pref}(\xi) \Delta \mathbf{pref}(\eta)) \cup (\mathbf{infix}(\xi) \Delta \mathbf{infix}(\eta))\} \\ \rho_\infty(\xi, \eta) &:= \sup\{r^{1-|w|} : w \in (\mathbf{pref}(\xi) \Delta \mathbf{pref}(\eta)) \cup (\mathbf{infix}^\infty(\xi) \Delta \mathbf{infix}^\infty(\eta))\}\end{aligned}$$

These metrics respect except for the length of the shortest non-common prefix of  $\xi$  and  $\eta$  also the length of the shortest non-common subword (non-common subword occurring infinitely often). Thus

$$\rho_I(\xi, \eta) \geq \rho(\xi, \eta) \text{ and } \rho_\infty(\xi, \eta) \geq \rho(\xi, \eta), \quad (2)$$

$$\rho_I(\xi, \eta) = \max\{\rho(\xi, \eta), \sup\{r^{1-|u|} : u \in \mathbf{infix}(\xi) \Delta \mathbf{infix}(\eta)\}\}, \text{ and} \quad (3)$$

$$\rho_\infty(\xi, \eta) = \max\{\rho(\xi, \eta), \sup\{r^{1-|u|} : u \in \mathbf{infix}^\infty(\xi) \Delta \mathbf{infix}^\infty(\eta)\}\}. \quad (4)$$

Similar to the case of  $\rho$  one can verify that  $\rho_I$  and  $\rho_\infty$  satisfy the ultra-metric inequality. Therefore, balls in the metric spaces  $(X^\omega, \rho_I)$  and  $(X^\omega, \rho_\infty)$  are simultaneously open and closed. Moreover, Eq. (2) shows that both topologies refine the CANTOR topology of  $X^\omega$ , that is,  $\omega$ -languages open (closed) in CANTOR topology are likewise open (closed, respectively) in both spaces  $(X^\omega, \rho_I)$  and  $(X^\omega, \rho_\infty)$ .

### 4.1 Shift-invariance

We call a metric space  $(X^\omega, \rho')$  *shift invariant* if for every open set  $E \subseteq X^\omega$  and every word  $w \in X^*$  the sets  $w \cdot E$  and  $E/w$  are also open. In this part we show that the metric spaces  $(X^\omega, \rho_\infty)$  and  $(X^\omega, \rho_I)$  are shift-invariant. According to Corollary 2 of [SS10] this property guarantees that the closure of a finite-state  $\omega$ -language is again finite-state (cf. the stronger Property 2.4 for the CANTOR topology).

To this end we derive some simple properties of the metrics.

**Lemma 1** *Let  $u \in X^*$  and  $v, w \in X^m$ . Then*

$$\rho_\infty(u \cdot \xi, u \cdot \eta) \leq \rho_\infty(\xi, \eta), \quad (5)$$

$$\rho_\infty(\xi, \eta) \leq r^m \cdot \rho_\infty(w \cdot \xi, v \cdot \eta), \quad (6)$$

$$\rho_I(u \cdot \xi, u \cdot \eta) \leq \rho_I(\xi, \eta), \text{ and} \quad (7)$$

$$\rho_I(\xi, \eta) \leq r^m \cdot \rho_I(w \cdot \xi, v \cdot \eta). \quad (8)$$

*Proof.* All inequalities are trivially satisfied if  $\xi = \eta$ . So, in the following, we may assume  $\xi \neq \eta$ .

As  $\text{infix}^\infty(\xi) = \text{infix}^\infty(u \cdot \xi)$ , Eqs. (5) and (6) follow from Eq. (4) and the respective properties of the metric  $\rho$  of the CANTOR topology  $\rho(u \cdot \xi, u \cdot \eta) \leq \rho(\xi, \eta)$  and  $\rho(w \cdot \xi, v \cdot \eta) \geq \rho(w \cdot \xi, w \cdot \eta) = r^{-|w|} \cdot \rho(\xi, \eta)$ .

Let  $\rho_I(\xi, \eta) = r^{-n}$ , that is,  $\mathbf{infix}(\xi) \cap X^n = \mathbf{infix}(\eta) \cap X^n$  and  $w \sqsubset \xi$  and  $w \sqsubset \eta$  for some  $w \in X^n$ . Then, obviously,  $v \sqsubset u \cdot \xi$  and  $v \sqsubset u \cdot \eta$  for some  $v \in X^n$ . Moreover,  $\mathbf{infix}(u \cdot \xi) \cap X^n = (\mathbf{infix}(u \cdot w) \cap X^n) \cup (\mathbf{infix}(\xi) \cap X^n) = \mathbf{infix}(u \cdot \eta) \cap X^n$ . This proves Eq. (7).

If  $w \neq v$  then in view of  $\rho(w \cdot \xi, v \cdot \eta) \geq r^{-(m-1)}$ , Eq. (8) is obvious. Let  $w = v$  and  $\rho_I(\xi, \eta) = r^{-n}$  for some  $n \in \mathbb{N}$ . We have to show that  $\rho_I(w \cdot \xi, w \cdot \eta) \geq r^{-(n+m)}$ .

If  $\rho(\xi, \eta) = r^{-n}$  then  $\rho(w \cdot \xi, w \cdot \eta) = r^{-(n+m)}$  and Eq. (3) proves  $\rho_I(w \cdot \xi, w \cdot \eta) \geq r^{-(n+m)}$ .

If  $\rho(\xi, \eta) < r^{-n}$  in view of  $\rho_I(\xi, \eta) = r^{-n}$  we have  $(\mathbf{infix}(\xi) \Delta \mathbf{infix}(\eta)) \cap X^{n+1} \neq \emptyset$ , that is,  $u \in (\mathbf{infix}(\xi) \Delta \mathbf{infix}(\eta)) \cap X^{n+1}$  for some  $u \in \mathbf{infix}(\xi)$ , say. Now, it suffices to show  $(\mathbf{infix}(w\xi) \Delta \mathbf{infix}(w\eta)) \cap X^{n+m+1} \neq \emptyset$ .

Assume  $v'u \notin \mathbf{infix}(w\xi) \Delta \mathbf{infix}(w\eta)$  for all  $v' \in X^m$ . Then  $u \in \mathbf{infix}(\xi)$  implies  $v'u \in \mathbf{infix}(w\xi) \cap \mathbf{infix}(w\eta)$  for some  $v' \in X^m$ . Since  $|w| = |v'| = m$ , we have  $u \in \mathbf{infix}(\eta)$ , a contradiction.  $\square$

As a consequence we obtain our result.

**Corollary 1** *The topologies  $(X^\omega, \rho_I)$  and  $(X^\omega, \rho_\infty)$  are shift invariant.*

*Proof.* We use the fact that, in view of Lemma 1, the mappings  $\Phi_u$  and  $\Phi_m$  defined by  $\Phi_u(\xi) := u \cdot \xi$  and  $\Phi_m(w \cdot \xi) := \xi$  for  $w \in X^m$  are continuous w.r.t. the metrics  $\rho_I$  and  $\rho_\infty$ , respectively.

Thus, if  $F \subseteq X^\omega$  is open in  $(X^\omega, \rho_I)$  or  $(X^\omega, \rho_\infty)$  then  $\Phi_u^{-1}(F) = F/u$  and, for  $m = |w|$ , also  $w \cdot F = \Phi_m^{-1}(F) \cap w \cdot X^\omega$  are open sets.  $\square$

## 4.2 Balls in $(X^\omega, \rho_I)$ and $(X^\omega, \rho_\infty)$

Denote by  $K_I(\xi, r^{-n})$  and  $K_\infty(\xi, r^{-n})$  the open balls<sup>5</sup> of radius  $r^{-n}$  around  $\xi$  in the spaces  $(X^\omega, \rho_I)$  and  $(X^\omega, \rho_\infty)$ , respectively. For  $w \sqsubset \xi$  with  $|w| = n+1$  and  $W := X^{n+1} \cap \mathbf{infix}(\xi)$ ,  $V := X^{n+1} \cap \text{infix}^\infty(\xi)$ ,  $\overline{W} := X^{n+1} \setminus \mathbf{infix}(\xi)$  and  $\overline{V} := X^{n+1} \setminus \text{infix}^\infty(\xi)$  we obtain the following description of balls via regular  $\omega$ -languages.

$$K_I(\xi, r^{-n}) = w \cdot X^\omega \cap \bigcap_{u \in W} X^* \cdot u \cdot X^\omega \setminus \bigcup_{u \in \overline{W}} X^* \cdot u \cdot X^\omega, \text{ and} \quad (9)$$

$$K_\infty(\xi, r^{-n}) = w \cdot X^\omega \cap X^* \cdot \left( \left( \prod_{u \in V} X^* \cdot u \right)^\omega \setminus \bigcup_{u \in \overline{V}} X^* \cdot u \cdot X^\omega \right). \quad (10)$$

<sup>5</sup>Since  $\rho_I$  and  $\rho_\infty$  satisfy the ultra-metric inequality, they are also closed balls of radius  $r^{-(n+1)}$ .

In Eq.(10) the order of the words  $u \in V$  can be arbitrarily chosen. In particular, Eqs. (9) and (10) show that balls in  $(X^\omega, \rho_I)$  and  $(X^\omega, \rho_\infty)$  are regular  $\omega$ -languages. Thus every non-empty open subset in each of the spaces contains an ultimately periodic  $\omega$ -word.

An immediate consequence of the representations in Eqs. (9) and (10) is the following relation between the space  $(X^\omega, \rho_I)$  and the CANTOR space  $(X^\omega, \rho)$ .

**Lemma 2** 1. Every ball  $K_I(\xi, r^{-n})$  is a Boolean combination of regular  $\omega$ -languages open in  $(X^\omega, \rho)$ , therefore, simultaneously an  $\mathbf{F}_\sigma$ - and a  $\mathbf{G}_\delta$ -set in the CANTOR topology.

2. Every open set in  $(X^\omega, \rho_I)$  is an  $\mathbf{F}_\sigma$ -set in CANTOR topology.

*Proof.* 1. It is well-known know that open sets in a metric space are simultaneously  $\mathbf{F}_\sigma$ - and  $\mathbf{G}_\delta$ -sets. Then, according to Property 1, the set  $K_I(\xi, r^{-n})$  is simultaneously an  $\mathbf{F}_\sigma$ - and  $\mathbf{G}_\delta$ -set in the CANTOR topology.

2. is a consequence of 1 and the fact that there are only countably many open balls in  $(X^\omega, \rho_I)$ .  $\square$

Eqs. (9) and (10) and Lemma 2 show a connection between certain regular  $\omega$ -languages and the open sets in  $(X^\omega, \rho_I)$ . It would be interesting if we could characterise some regular  $\omega$ -languages open in  $(X^\omega, \rho_I)$  using CANTOR topology. The next example considering the simple case of closed sets, however, shows that not every regular  $\omega$ -language closed in CANTOR topology is open in  $(X^\omega, \rho_I)$ .

**Example 2 ([Ho14])** Consider the regular  $\omega$ -language  $F = \{1, 00\}^\omega \subseteq \{0, 1\}^\omega$  which is closed in the CANTOR topology. Assume  $F$  to be open in  $(X^\omega, \rho_I)$ . Then  $\eta = \prod_{i \in \mathbb{N}} 10^{2i} \in F$  and, therefore,  $K_I(\eta, r^{-n}) \subseteq F$  for some  $n \in \mathbb{N}$ ,  $n \geq 1$ .

Consider  $\xi = \prod_{i=0}^n 10^{2i} \cdot \prod_{i=2n+1}^\infty 10^i \notin F$ . Then we have  $\prod_{i=0}^n 10^{2i} \sqsubset \eta$ ,  $\prod_{i=0}^n 10^{2i} \sqsubset \xi$  and, moreover,

$$\begin{aligned} \mathbf{infix}(\xi) \cap \{0, 1\}^{2n} &= (\mathbf{infix}(\prod_{i=0}^n 10^{2i}) \cup 0^* \cdot 1 \cdot 0^* \cup 0^*) \cap \{0, 1\}^{2n} \\ &= \mathbf{infix}(\eta) \cap \{0, 1\}^{2n}. \end{aligned}$$

It follows  $\rho_I(\xi, \eta) \leq r^{-2n}$ , that is,  $\xi \in K_I(\eta, r^{-n}) \subseteq \{1, 00\}^\omega$ , a contradiction.  $\square$

Using the Morse-Hedlund Theorem (cf. also the proof of Theorem 1.3.13 of [Lo02]) one obtains special representations of small balls containing ultimately periodic  $\omega$ -words. To this end we derive the following lemma.

**Lemma 3** Let  $w, u \in X^*$ ,  $u \neq e$  and  $\xi \in X^\omega$ . Then  $w \cdot u \sqsubset \xi$  and  $\mathbf{infix}(\xi) \cap X^{|w \cdot u|} = \mathbf{infix}(w \cdot u^\omega) \cap X^{|w \cdot u|}$  imply  $\xi = w \cdot u^\omega$ .

*Proof.* First observe that  $|\mathbf{infix}(w \cdot u^\omega) \cap X^{|w \cdot u|}| = |\mathbf{infix}(w \cdot u^\omega) \cap X^{|w \cdot u|+1}|$ . Thus, for every  $v \in \mathbf{infix}(w \cdot u^\omega) \cap X^{|w \cdot u|}$ , there is a unique  $v' \in \mathbf{infix}(w \cdot u^\omega) \cap X^{|w \cdot u|}$  such that  $v \sqsubset a \cdot v'$  for some  $a \in X$ . Consequently, the  $\omega$ -word  $\xi \in X^\omega$  with  $w \cdot u \sqsubset \xi$  and  $\mathbf{infix}(\xi) \cap X^{|w \cdot u|} = \mathbf{infix}(w \cdot u^\omega) \cap X^{|w \cdot u|}$  is uniquely specified.

**Lemma 4** *Let  $w \cdot u^\omega \in X^\omega$  where  $|w| \leq |u|$  and let  $m > |w| + |u|$  and  $n > |u|$ . Then*

$$K_I(w \cdot u^\omega, r^{-m}) = \{w \cdot u^\omega\}, \text{ and} \quad (11)$$

$$K_\infty(w \cdot u^\omega, r^{-n}) = w' \cdot X^* \cdot u^\omega \text{ where } w' \sqsubset w \cdot u \text{ and } |w'| = n. \quad (12)$$

*Proof.* Every  $\xi \in K_I(w \cdot u^\omega, r^{-m})$  satisfies  $w \cdot u \sqsubset \xi$  and  $\mathbf{infix}(\xi) \cap X^m = \mathbf{infix}(w \cdot u^\omega) \cap X^m$ , and the assertion of Eq. (11) follows from Lemma 3.

If  $\xi \in K_\infty(w \cdot u^\omega, r^{-n})$  then there is a tail  $\xi'$  of  $\xi$  such that  $u \sqsubset \xi'$  and  $\mathbf{infix}^\infty(\xi) \cap X^n = \mathbf{infix}(\xi') \cap X^n = \mathbf{infix}(u^\omega) \cap X^n$  whence, again by Lemma 3,  $\xi' = u^\omega$ .  $\square$

This allows us to state the following property concerning isolated points<sup>6</sup> in the spaces  $(X^\omega, \rho_I)$  and  $(X^\omega, \rho_\infty)$ . The additional Item 3 in connection with Lemma 2.2 shows a further difference between both spaces.

**Corollary 2** 1. *The set of isolated points of the space  $(X^\omega, \rho_I)$  is Ult.*

2. *The space  $(X^\omega, \rho_\infty)$  has no isolated points and all sets of the form  $X^* \cdot u^\omega$  are simultaneously closed and open.*

3. *In the space  $(X^\omega, \rho_\infty)$  there are open sets which are not  $\mathbf{F}_\sigma$ -sets in CANTOR topology.*

*Proof.* Since every non-empty open subset of  $(X^\omega, \rho_I)$  and also  $(X^\omega, \rho_\infty)$  contains an ultimately periodic  $\omega$ -word, every isolated point has to be ultimately periodic. Now Eq. (11) shows that every  $w \cdot u^\omega$  is an isolated point in  $(X^\omega, \rho_I)$ , and Eq. (12) proves that  $(X^\omega, \rho_\infty)$  has no isolated points. The remaining part of Item 2 follows from Eq. (12) and  $X^* \cdot u^\omega = \bigcup_{w \in X^n} w \cdot X^* \cdot u^\omega$ .

Finally, it is known that  $X^\omega \setminus X^* \cdot u^\omega$  is not an  $\mathbf{F}_\sigma$ -set in CANTOR topology (cf. Example 1).  $\square$

### 4.3 Non-preservation of regular $\omega$ -languages

In this section we investigate whether similar to the CANTOR topology the closure of a finite-state  $\omega$ -language is always regular in the spaces  $(X^\omega, \rho_I)$  and  $(X^\omega, \rho_\infty)$ .

<sup>6</sup>A point  $\xi$  is referred to as *isolated* if  $\rho'(\xi, \eta) \geq \varepsilon_\xi$  for all  $\eta \neq \xi$ . Here the distance  $\varepsilon_\xi > 0$  may depend on  $\xi$ .

In contrast to the CANTOR topology it is, however, not true that the closure of finite-state  $\omega$ -languages are regular. We can even show that in both spaces  $(X^\omega, \rho_I)$  and  $(X^\omega, \rho_\infty)$  there are regular  $\omega$ -languages with non-regular closures.

Since we do not have a characterisation like the Property 1.5 for the closures  $\mathcal{C}_I$  and  $\mathcal{C}_\infty$  in the spaces  $(X^\omega, \rho_I)$  and  $(X^\omega, \rho_\infty)$ , respectively, we circumvent this obstacle by presenting examples where the closure  $\mathcal{C}_I(F)$  or  $\mathcal{C}_\infty(F)$  of a regular  $\omega$ -language  $F$  is shown to be larger than  $F$  but does not contain more ultimately periodic  $\omega$ -words than  $F$ . In view of Theorem 1 this implies that the closures cannot be regular  $\omega$ -languages.

For the closure  $\mathcal{C}_I$  we use that, according to Example 1 the  $\omega$ -language  $\{0, 1\}^* \cdot 0^\omega$  is no  $\mathbf{G}_\delta$ -set in the CANTOR topology, thus in view of Lemma 2.2 not closed in  $(X^\omega, \rho_I)$ .

**Example 3** We show that  $\mathcal{C}_I(\{0, 1\}^* \cdot 0^\omega) \cap \text{Ult} = \{0, 1\}^* \cdot 0^\omega$ . Let  $w \cdot u^\omega \notin \{0, 1\}^* \cdot 0^\omega$ . Then  $u \notin \{0\}^*$  and  $0^{|w \cdot u|} \notin \mathbf{infix}(w \cdot u^\omega)$ . Now Eq. (9) yields  $K(w \cdot u^\omega, r^{-|w \cdot u|}) \cap X^* \cdot 0^{|w \cdot u|} \cdot X^\omega = \emptyset$ . Thus  $\rho_I(w \cdot u^\omega, v \cdot 0^\omega) \geq r^{-|w \cdot u|}$  for all  $v \in X^*$  whence  $w \cdot u^\omega \notin \mathcal{C}_I(\{0, 1\}^* \cdot 0^\omega)$ . The other inclusion being trivial.

Assume  $\mathcal{C}_I(\{0, 1\}^* \cdot 0^\omega)$  were a regular  $\omega$ -language. Then Theorem 1 implies  $\mathcal{C}_A(\{0, 1\}^* \cdot 0^\omega) = \{0, 1\}^* \cdot 0^\omega$ , that is,  $\{0, 1\}^* \cdot 0^\omega$  is closed in  $(X^\omega, \rho_I)$ , a contradiction to Lemma 2.2.  $\square$

Since  $\{0, 1\}^* \cdot 0^\omega$  is closed in  $(X^\omega, \rho_\infty)$ , we cannot use this  $\omega$ -language in that case.

**Example 4**  $\xi_i \cap X^n = \mathbf{infix}^\infty(\zeta) \cap X^n$  for  $n \leq 2i + 1$ . This implies  $\rho_\infty(\xi_i, \zeta) \leq r^{-2i}$ , that is,  $\lim_{i \rightarrow \infty} \xi_i = \zeta \in \mathcal{C}_\infty(F)$  in  $(X^\omega, \rho_\infty)$ .  $\square$

## 5 Completeness and Compactness

Here we show that the spaces  $(X^\omega, \rho_I)$  and  $(X^\omega, \rho_\infty)$  are neither complete nor compact.

To show that they are not complete we consider the sequence  $(\xi_i)_{i \in \mathbb{N}}$  where  $\xi_i := \prod_{j=i}^\infty 0^j 1$ . This sequence converges in CANTOR topology to the limit point  $0^\omega$ . Since  $(X^\omega, \rho_I)$  and  $(X^\omega, \rho_\infty)$  refine  $(X^\omega, \rho)$ , the limit points, if they exist, should be the same. But  $\mathbf{infix}(\xi_i)$  and  $\mathbf{infix}^\infty(\xi_i)$  both contain the word 1 which is not in  $\mathbf{infix}(0^\omega) = \mathbf{infix}^\infty(0^\omega)$ . Thus  $\rho_I(\xi_i, 0^\omega) = \rho_\infty(\xi_i, 0^\omega) = 1$ .

It remains to show that the sequence  $(\xi_i)_{i \in \mathbb{N}}$  fulfils the CAUCHY property. To this end we observe that for  $j \geq i$  we have  $0^i \sqsubset \xi_j$  and  $\mathbf{infix}(\xi_j) \cap X^i = \mathbf{infix}^\infty(\xi_j) \cap X^i = \{0^i\} \cup \{0^m 10^{i-m-1} : 0 \leq m < i\}$ . Thus  $\rho_I(\xi_j, \xi_k) \leq r^{-i}$  and also  $\rho_\infty(\xi_j, \xi_k) \leq r^{-i}$  for  $j, k \geq i$ .

In general it holds that no topology refining the CANTOR topology is compact. A proof uses Corollary 3.1.14 in [En77]. Here we provide the more illustrative and seemingly stronger examples of partitions of the whole space  $X^\omega$  into infinitely many open subsets.

**Example 5** Let  $X = \{0, 1\}$ . Then the sets  $0^i 1 \cdot X^\omega$  for  $i \in \mathbb{N}$  are open in the CANTOR topology, hence open in  $(X^\omega, \rho_I)$  and according to Corollary 2.1 the set  $\{0^\omega\}$  is also open  $(X^\omega, \rho_I)$ .

Then  $\{\{0^\omega\} \cup \{0^i 1 \cdot X^\omega : i \in \mathbb{N}\}\}$  is a partition of  $X^\omega$  into sets open in  $(X^\omega, \rho_I)$ .

**Example 6** Let  $X = \{0, 1\}$ . Then the sets  $0^i 1 \cdot X^\omega$  for  $i \in \mathbb{N}$  are open in the CANTOR topology, hence open in  $(X^\omega, \rho_\infty)$  and according to Corollary 2.2 the set  $X^* \cdot 0^\omega$  is open and closed in  $(X^\omega, \rho_\infty)$ .

Then  $\{X^* \cdot 0^\omega\} \cup \{0^i 1 \cdot X^\omega \setminus X^* \cdot 0^\omega : i \in \mathbb{N}\}$  is a partition of  $X^\omega$  into sets open in  $(X^\omega, \rho_\infty)$ .

## 6 Subword Complexity

In Section 4 we mentioned that regular  $\omega$ -languages are closely related to the (asymptotic) subword complexity of infinite words. Adapting the metrics  $\rho_I$  and  $\rho_\infty$  to subwords we may draw some connections to the level sets  $F_\gamma^{(\tau)}$  of the asymptotic subword complexity (see [St93, St12]).

First we introduce the concept of asymptotic subword complexity.

**Definition 2 (Asymptotic subword complexity)**

$$\tau(\xi) := \lim_{n \rightarrow \infty} \frac{\log_{|X|} |\mathbf{infix}(\xi) \cap X^n|}{n}$$

Using the inequality  $|\mathbf{infix}(\xi) \cap X^{n+m}| \leq |\mathbf{infix}(\xi) \cap X^n| \cdot |\mathbf{infix}(\xi) \cap X^m|$  it is easy to see that the limit in Definition 2 exists and

$$\tau(\xi) = \inf \left\{ \frac{\log_{|X|} |\mathbf{infix}(\xi) \cap X^n|}{n} : n \in \mathbb{N} \wedge n \geq 1 \right\}. \quad (13)$$

Eq. (5.2) of [St93] shows that in Definition 2 and Eq. (13) one can replace the term  $\mathbf{infix}(\xi)$  by  $\mathbf{infix}^\infty(\xi)$ .

Let, for  $0 < \gamma \leq 1$ ,  $F_\gamma^{(\tau)} := \{\xi : \xi \in X^\omega \wedge \tau(\xi) < \gamma\}$  be the *lower level sets* of the asymptotic subword complexity. For  $\gamma = 0$  we set  $F_0^{(\tau)} := \text{Ult}$  (instead of  $F_0^{(\tau)} = \emptyset$ ). We want to show that these sets are open in  $(X^\omega, \rho_I)$  and  $(X^\omega, \rho_\infty)$ . As a preparatory result we derive the subsequent Lemma 5.

Let  $E_n(\xi) := \{\eta : \mathbf{infix}(\eta) \cap X^n \subseteq \mathbf{infix}(\xi)\}$  and  $E'_n(\xi) := \{\eta : \mathbf{infix}^\infty(\eta) \cap X^n \subseteq \mathbf{infix}^\infty(\xi)\}$  be the sets of  $\omega$ -words having only infixes or infixes occurring infinitely often of length  $n$  of  $\xi$ , respectively. These sets can be equivalently described as

$$E_n(\xi) = X^\omega \setminus X^* \cdot (X^n \setminus \mathbf{infix}(\xi)) \cdot X^\omega \quad \text{and}$$

$$E'_n(\xi) = X^* \cdot (X^\omega \setminus X^* \cdot (X^n \setminus \mathbf{infix}^\infty(\xi)) \cdot X^\omega), \quad \text{respectively}$$

which resembles in some sense the characterisation of open balls in Eqs. (9) and (10). In fact, it appears that the sets  $E_n(\xi)$  and  $E'_n(\xi)$  are open in the respective spaces  $(X^\omega, \rho_I)$  and  $(X^\omega, \rho_\infty)$ .

**Lemma 5** *Let  $\xi \in X^\omega$ . Then  $\xi \in E_n(\xi) \cap E'_n(\xi)$ , the set  $E_n(\xi)$  is open in  $(X^\omega, \rho_I)$  and the set  $E'_n(\xi)$  is open in  $(X^\omega, \rho_\infty)$ .*

*Proof.* The first assertion is obvious. For a proof of the second one we show that  $\eta \in E_n(\xi)$  implies that the ball  $K_I(\eta, r^{-n})$  is contained in  $E_n(\xi)$ .

Let  $\eta \in E_n(\xi)$  and  $\zeta \in K_I(\eta, r^{-n})$ . Then,  $\rho_I(\eta, \zeta) < r^{-n}$ , that is, in particular,  $\mathbf{infix}(\eta) \cap X^n = \mathbf{infix}(\zeta) \cap X^n$ , whence  $\zeta \in E_n(\xi)$

The proof for  $E'_n(\xi)$  is similar.  $\square$

This much preparation enables us to show that the level sets are open sets.

**Theorem 3** *Let  $0 \leq \gamma \leq 1$ . Then the sets  $F_\gamma^{(\tau)}$  are open in  $(X^\omega, \rho_I)$  and  $(X^\omega, \rho_\infty)$ .*

*Proof.* For  $\gamma = 0$  we have  $F_\gamma^{(\tau)} = \text{Ult}$  which is, according to Corollary 2, open as well in  $(X^\omega, \rho_I)$  as in  $(X^\omega, \rho_\infty)$ .

Let  $\gamma > 0$  and  $\tau(\xi) < \gamma$ . We show that then  $E_n(\xi) \subseteq F_\gamma^{(\tau)}$  and  $E'_n(\xi) \subseteq F_\gamma^{(\tau)}$  for some  $n \in \mathbb{N}$ . Together with Lemma 5 this shows that  $F_\gamma^{(\tau)}$  contains, with every  $\xi$ , open sets containing this  $\xi$ .

If  $\tau(\xi) < \gamma$  then in view of Eq. (13) for some  $n \in \mathbb{N}$  we have  $\frac{\log_{|X|} |\mathbf{infix}(\xi) \cap X^n|}{n} < \gamma$ . Then for every  $\eta \in E_n(\xi)$  it holds  $\tau(\eta) \leq \frac{\log_{|X|} |\mathbf{infix}(\eta) \cap X^n|}{n} < \gamma$  and, consequently,  $E_n(\xi) \subseteq F_\gamma^{(\tau)}$ .

The proof for  $(X^\omega, \rho_\infty)$  is similar using  $\mathbf{infix}^{(\infty)}$  instead of  $\mathbf{infix}$  and the respective modification of Eq. (13) whose validity was mentioned above.  $\square$

The proof shows also that  $\xi \in F_\gamma^{(\tau)}$  implies that  $X^\omega \setminus X^* \cdot (X^n \setminus \mathbf{infix}(\xi)) \cdot X^\omega \subseteq F_\gamma^{(\tau)}$  for some  $n > 0$ . Thus  $F_\gamma^{(\tau)}$  is a countable union of regular  $\omega$ -languages closed in CANTOR topology, hence an  $\mathbf{F}_\sigma$ -set in CANTOR topology. The sets  $F_\gamma^{(\tau)}$  are finite-state<sup>7</sup> non-regular  $\omega$ -languages because their complement  $X^\omega \setminus F_\gamma^{(\tau)}$  is non-empty and does not contain any ultimately periodic  $\omega$ -word. Thus, in view of Theorem 2, they are not  $\mathbf{G}_\delta$ -sets in CANTOR-space and they are examples of

<sup>7</sup>In particular, they satisfy  $F_\gamma^{(\tau)} / w = F_\gamma^{(\tau)}$  for  $w \in X^*$ .

sets open in  $(X^\omega, \rho_I)$  and  $(X^\omega, \rho_\infty)$  which are non-regular  $\mathbf{F}_\sigma$ -sets in CANTOR-space.

## References

- [Bü60] Büchi, J.R., On a decision method in restricted second order arithmetic. Proc. 1960 Int. Congr. for Logic, Stanford Univ. Press, Stanford 1962, 1–11.
- [CM03] Calude, C.S., Marcus, S. and Staiger, L., A topological characterization of random sequences, Inform. Process. Lett., Vol. 88 (2003), 245-250.
- [CJ09] Calude, C.S., Jürgensen H., and Staiger, L., Topology on words, Theoret. Comput. Sci. 410 (2009), 2323-2335.
- [DK09] Diekert, V. and Kufleitner, M., Fragments of first-order logic over infinite words, In: Proc. STACS'09 (S. Albers and J.-Y. Marion, eds.), Schloss Dagstuhl - Leibniz-Zentrum für Informatik, Germany, 2009, 325–336.
- [En77] Engelking, R., *General Topology*, Państwowe wydawnictwo naukowe, Warszawa 1977.
- [FS01] Fernau, H. and Staiger, L., Iterated function systems and control languages, Inform. and Comput., Vol. 168 (2001), 125-143.
- [Ho14] Hoffmann, S., Metriken zur Verfeinerung des Cantor-Raumes auf  $X^\omega$ . Diploma thesis, Martin-Luther-Universität Halle-Wittenberg, 2014.
- [La69] Landweber, L.H., Decision problems for  $\omega$ -automata, Math. Syst. Theory 3(1969) 4, 376–384.
- [Lo02] Lothaire, M.. *Algebraic Combinatorics on Words*, Cambridge University Press, Cambridge 2002.
- [PP04] Perrin, D. and Pin, J.-E.. *Infinite Words*, Elsevier, Amsterdam 2004.
- [Re86] Redziejowski, R.R., Infinite word languages and continuous mappings. Theoret. Comput. Sci. 43 (1986) 1, 59–79.
- [RS97] Rozenberg, G. and Salomaa, A. (eds.), *Handbook of Formal Languages*, Springer, Berlin (1997).
- [SS10] Schwarz, S. and Staiger, L. *Topologies refining the Cantor topology on  $X^\omega$* , in: IFIP Advances in Information and Communication Technology Vol. 323, Springer-Verlag, Berlin 2010, 271 - 285,

- [St83] Staiger, L., Finite-state  $\omega$ -languages, *J. Comput. Syst. Sci.* 27 (1983) 3, 434-448.
- [St87] Staiger, L., Sequential mappings of  $\omega$ -languages, *ITA* 21 (1987) 2, 147-173.
- [St93] Staiger, L., Kolmogorov complexity and Hausdorff dimension, *Inf. Comput.* 103 (1993) 2, 159-194.
- [St97] Staiger, L.,  $\omega$ -languages, In: [RS97], Vol. 3, 339–387.
- [St03] Staiger, L., Weighted finite automata and metrics in Cantor Space, *Journal of Automata, Languages and Combinatorics* 8 (2003) 2, 353-360.
- [St05] Staiger, L., Topologies for the set of disjunctive  $\omega$ -words, *Acta Cybern.* 17 (2005) 1, 43-51.
- [St12] Staiger, L., Asymptotic subword complexity, In: *Languages Alive* (H. Bordihn, M. Kutrib and B. Truthe, eds.), *Lect. Notes Comput. Sci.* 7000, Springer-Verlag, Heidelberg 2012, 236–245.
- [SW74] Staiger, L., and Wagner, K., Automatentheoretische und automatenfreie Charakterisierungen topologischer Klassen regulärer Folgenmengen, *Elektronische Informationsverarbeitung und Kybernetik*, 10 (1974) 7, 379-392.
- [Th90] Thomas, W., Automata on infinite objects, In: J. Van Leeuwen (Ed.), *Handbook of Theoretical Computer Science*, Vol. B, 133–191, Elsevier, Amsterdam, 1990.
- [Th97] Thomas, W., Languages, automata, and logic, In: [RS97], Vol. 3, 389–455.
- [Tr62] Trakhtenbrot, B.A., Finite automata and monadic second order logic, *Sibirsk. Mat. Ž.* 3 (1962), 103–131. (Russian; English translation: *AMS Transl.* 59 (1966), 23–55.)