LATVIJAS
UNIVERSITĀTE
ANNO 1919

UNIVERSITY OF LATVIA
FACULTY OF PHYSICS AND MATHEMATICS
DEPARTMENT OF MATHEMATICS

STRUCTURE OF RECURRENT WORDS: RESISTANCE AND MEASURE OF PROXIMITY

PhD Thesis

Author: **Raivis Bēts**

Student ID rb09224

Scientific advisor: Prof. Jānis Buls

RIGA 2016

ESF
EIROPAS SOCIĀLAIS FONDS

EIROPAS SAVIENĪBA

LATVIJAS UNIVERSITĀTE
ANNO 1919

IEGULDĪJUMS TAVĀ NĀKOTNĒ

This work has been supported by the European Social Fund within the two projects:
“Support for Doctoral Studies at University of Latvia” and
project No. 2013/0024/1DP/1.1.1.2.0/13/APIA/VIAA/045.

Anotācija

Disertācijā tiek piedāvāts aperiodiska gadījuma skaitļu filtrācijas ģenerators. Šajā ģeneratorā filtra virknes vietā tiek izmantots galīgi ģenerēts bi-ideāls. Disertācijā dota metode, kas dotajai periodiskai virknei konstruē bezgalīgi daudz tādus galīgi ģenerētus bi-ideālus, ka filtrācijas ģeneratora rezultējošā virkne ir aperiodiska. Pierādīta universālu bi-idealu eksistence. Tie ir tādi bi-ideāli, kuri filtrācijas rezultātā rada aperiodiskus vārdus, iedarbojoties uz visām netriviālajām, periodiskajām virknēm.

Disertācijā aplūkots un analizēts kombinatorisks nosacījums – labi sadalīto ieeju jeb WELLDOC īpašība. Analizēta WELLDOC īpašība ierobežotiem bi-ideāliem. Pierādīta galīgā alfabētā 1-ierobežota bi-ideāla eksistence, kas apmierina WELLDOC īpašību. Dots piemērs, kas ilustrē patvaļīga 1-ierobežota bi-ideāla iegūšanas procedūru. Tiek ieviesti pilnīgi ierobežoti bi-ideāli, kuriem tiek uzlikti nosacījumi uz to ģenerējošajām bāzēm. Darbā pierādīts, ka ierobežoti bi-ideāli ir lineāri rekurenti tad un tikai tad, ja tie ir pilnīgi ierobežoti.

Disertācijā tiek aplūkoti daļēji vārdi, t.i., vārdi, kas satur tā sauktos ”nezināmos” simbolus. Pierādīts, ka dotam galīgi ģenerētam bi-ideālam ir iespējams atrast ģenerējošo bāzi. Tāpat pierādīts, ka dotam bi-ideālam ir iespējams konstruēt tādu bāzi, kas ģenerē sākotnējo bi-ideālu bez pirmā burta. Disertācijā dots pierādījums, ka galīgi ģenerētam bi-ideālam ar galīgu skaitu ”nezināmo” simbolu tos ir iespējams aizpildīt. Pierādīts, ka divi nereducējami galīgi ģenerēti bi-ideāli satur bezgalīgi daudz ieejas, kurās to simboli ir atšķirīgi. Ir aprakstīts un dots dažādu bezgalīgu vārdu klašu mērs. Disertācijā piedāvāta nestriktas metrikas konstrukcija uz bezgalīgu vārdu kopas. Doti piemēri, kas parāda nestriktas metrikas priekšrocības, salīdzinot to ar standarta metrikām vārdu kombinatorikā.

Atslēgas vārdi: Aperiodisks filtrācijas ģenerators, galīgi ģenerēts bi-ideāls, ierobežots bi-ideāls, WELLDOC īpašība, daļēji vārdi, nestrikta metrika.

Abstract

The thesis presents a non-periodic random number generator based on the shrinking generator. The A-sequence is still generated using an LFSR, but the S-sequence is replaced by a finitely generated bi-ideal — an aperiodic sequence. A method for the construction of an infinite number of finitely generated bi-ideals from a given A-sequence, such that the resulting sequence of the shrinking generator is aperiodic is shown. The existence of what we call universal finitely generated bi-ideals that produce aperiodic words when used as the S-sequence of a shrinking generator for all non-trivial periodic A-sequences is proved.

A combinatorial condition called well distributed occurrences, or WELLDOC for short, has been explored in the thesis. The WELLDOC property for bounded bi-ideals is analysed in the thesis. The existence of a 1-bounded bi-ideal over the finite alphabet that satisfies the WELLDOC property has been proved in the thesis. An example of obtaining and achieving arbitrary 1-bounded bi-ideal with the WELLDOC property is given. The notion of completely bounded bi-ideals by imposing a restriction on their generating base sequences is introduced. We prove that a bounded bi-ideal is linearly recurrent if and only if it is completely bounded.

The thesis explores partial words, i.e., words which contain so called ”do not know” symbols. The proof that for a given finitely generated bi-ideal sequence possibility of finding the basis is given. It is proved that for a given bi-ideal it is possible to construct the basis for the same bi-ideal without the first letter. The thesis also contains the proof of possibility to fill the finite number of holes for a given finitely generated bi-ideal. The fact that two irreducible finitely generated bi-ideals have infinitely many differ symbols is proved. Measures of different classes of infinite words are given. The thesis introduces a new metric on the set of infinite words. Construction of a fuzzy metric on the set of infinite words is given. Also, examples that show advantages comparing fuzzy metric with standart metrics in combinatorics on words is given.

Keywords: Aperiodic shrinking generator, finitely generated bi-ideal, bounded bi-ideal, WELLDOC property, partial words, fuzzy metric.

Acknowledgements

First and foremost I want to offer my thankfulness to my scientific supervisor prof. Jānis Buls for introducing me to this very interesting mathematical field – Combinatorics on Words 7 years ago. His help, advice and kindness all these years was very important for me.

Secondly, I want to say thanks to all the participants of the seminar ”Combinatorics on words” – Inese, Edmunds and Līga for the worthwhile debates and useful presentations.

The financial support of the University of Latvia and the European Social Fund provided the chance of doing what I enjoy for these several past years.

Finally, I offer my warmest gratitudes to my family and friends, especially to my mother and Evita, for patience, understanding and support I have received from them. Finally, thanks to my choir ”Aura” and its conductor Edgars, who helped to clear my mind, when I was feeling stressed through all these years.

Contents

Notions

$\mathbb{N}$	the set of all non-negative integers,
$\mathbb{N}_+$	the set of all positive integers,
Σ_n	the set $\{0, 1, 2, \ldots, n-1\}$ for some $n \in \mathbb{N}_+$,
$\overline{i, j}$	the set $\{i, i+1, i+2, \ldots, j-1, j\}$ for some $i, j \in \mathbb{N}$, $i \leq j$,
$\gcd(n, k)$	the greatest common divisor of numbers n and k,
A^*	the set of all finite words over an alphabet A,
A^+	the set of all finite non-empty words over an alphabet A,
A^ω	the set of all infinite words over an alphabet A,
λ	the empty word,
u^n	a finite word $\underbrace{uu\cdots u}_{n}$, where $n \in \mathbb{N}_+$ and u is a finite word,
u^ω	the infinite word $uu\cdots u\cdots$, where u is a finite non-empty word,
$\lvert u\rvert$	the length of the finite word u,
$\lvert u\rvert_a$	the number of occurrences of the letter a in word the u,
$\lvert u\rvert_v$	the number of occurrences of the word v in word the u,
$alph(u)$	the set of all letters occurring in the word u,
$x[i, j]$	the factor of a word x starting in position i and ending in position j, where $i < j$, in other words, $x_i x_{i+1} \cdots x_{j-1} x_j$,
$x[i, j)$	the factor of a word x starting in position i and ending in position $j-1$, where $i+1 < j$, in other words, $x_i x_{i+1} \cdots x_{j-1}$,
$pref_n u$	prefix of length n of a word u.

Pref(w)	the set of all prefixes of a word x,
F(w)	the set of all factors of a word x,
Suff(w)	the set of all suffixes of a word x,
$\mathcal{R}_{x,u}$	the set of all return words to u of x,
$\langle u_0, u_1, \ldots, u_{m-1}\rangle$	a basis of a finitely generated bi-ideal,
$\langle u_0^{(n)}, u_1^{(n)}, \ldots, u_{m-1}^{(n)}\rangle$	a basis of a finitely generated bi-ideal that is obtained when the basis $\langle u_0, u_1, \ldots, u_{m-1}\rangle$ is L-prolonged n times,
$(u_n^{(k)})_{n\geq 0}$	the basis sequence of a bi-ideal that is obtained when the basis sequence $(u_n)_{n\geq 0}$ is L-prolonged k times
$u_i^{(k)}$	the i-th element of the basis sequence that is obtained, when the initial basis sequence $(u_n)_{n\geq 0}$ is L-prolonged k times,
$(\lvert w\rvert_0, \lvert w\rvert_1, \ldots, \lvert w\rvert_{d-1})$	Parikh vector of a finite word w.

Introduction

Combinatorics on words is a relatively new field of discrete mathematics whose history started a bit more than a hundred years ago, when a Norwegian mathematician Thue published his first paper on repetition free words (Thue, 1906) and (Thue, 1912). It is considered as a starting point of research on words. He published his papers in his native country (Oslo University) and it became the reason his work remained unknown for a while although some of his results were rediscovered by other scientists.

Highly significant work was done by Morse and Hedlund (Morse and Hedlund, 1938), who founded the field of symbolic dynamics. The work of Morse got an intersection with Thue word and resulted in the Thue-Morse infinite word, which still has an impact on the field of combinatorics on words. This historical word also have a connection with such scientific fields as algebra, finite automata and others. Morse together with Hedlund introduced world to Sturmian sequences (Morse and Hedlund, 1940), which still have an enormous impact nowadays.

The biggest revulsion of the combinatorics on words started in the middle of last century when the theory of words was developed in Russia and France almost at the same time. A massive influence in Russia was made by Novikov and Adjan and it resulted in some important papers (Novikov, 1955) and (Adjan, 1979). The impact in France born from reserch of Schützenberger, who worked on theory of codes (Schützenberger, 1956). Almost ten years later he gave the insights in theory of context-free languages (Chomsky and Schützenberger, 1963) and factorizations of free monoids (Schützenberger, 1965).

At this time the ground of theory of words was settled, so more and more scientific results and theories ensued. This all lead down to the first book of the field, called Combinatorics on Words (Lothaire, 1983), which was written by a group of authors. This book turned topic of words a challenge on its own. 20 years later developements of words brought to the second book – Algebraic Combinatorics on Words (Lothaire, 2002) and this book repeated nothing from the first one.

Since 1997 the conference – WORDS, devoted entirely to combinatorics on words, has been created and every second year it brings many researchers of this challenging field to it. Nowadays combinatorics on words is connected to many other topics: algebra, probability theory, automata theory, biology, physics, algorithms and others, but it still has remained as a research topic of itself as well. For a good insight, summary of history and for new challenges we recommend (Berstel and Perrin, 2007), (Karhumäki, 2004) and (Berstel and Karhumäki, 2003).

Combinatorics on words is dealing with words – finite or infinite sequences of symbols (letters). The main object in this thesis are infinite words (or ω-words). We are dealing with a subclass of infinite words, so called recurrent words. Such words have a property that every factor occurs in them infinitely many times. We use an equivalent notion – a bi-ideal, which is an infinite word containing as prefixes all elements of a bi-ideal sequence (Coudrain and Schützenberger, 1966). A bi-ideal sequence is a sequence of words such that each next element of the sequence is at least twice as long as the previous element and contains the previous element both its prefix and suffix. Also, the words in a bi-ideal sequence are known as Zimin's words (Zimin, 1982) or sesquipowers (Simon, 1988). Buls and Lorencs investigated bi-ideals from different aspects. Regularities of periodicity in bi-ideals have been prospected in (Buls and Lorencs, 2006) and (Buls and Lorencs, 2008), but Lorencs and Cers tried and successfully solved the decision problem of finitely generated bi-ideals (Cers, 2010), (Cers, 2012) and (Lorencs, 2012).

As recurrent words and bi-ideals are describing the same class of infinite words (see e.g., (de Luca and Varricchio, 1999)), we choose to view this class as class of bi-ideals since they have a nice and useful structure. A lot of the proofs in this thesis use this structure of bi-ideals. As almost every word is a recurrent word, then bi-ideals have the same property. Bi-ideals are covering almost the whole class of infinite words, so they have a significant role in the class of infinite words. Mostly in this thesis we are dealing with subclasses of bi-ideals – class of bounded bi-ideals and class of finitely generated bi-ideals (which is subclass of bounded bi-ideals). While bounded bi-ideals have restriction on the length of base words, finitely generated bi-ideals are generated by a periodic basis.

Applications in cryptology

In this thesis, we propose a method to generate non-periodic pseudo-random number sequences based on the shrinking generator modification, which was introduced by Coppersmith et.al. in 1993 (see (Coppersmith et al., 1994)) and is still considered to be a secure pseudo-random number generator. Normally, a shrinking generator uses two pseudo-random bit-sequences produced by LFSR's (see, e.g., (Schneier and Sutherland, 1995)) from which the resulting pseudo-random sequence is obtained by taking the subsequence of one of the sequences (called the A-sequence) corresponding to the positions of ones in the other sequence (called the S-sequence).

In this thesis we show two approaches for generating non-periodic pseudo-random number sequences using our modified shrinking generator. Firstly, given a periodic A-sequence, we prove that any finitely generated bi-ideal that satisfies a simple condition can be used as the S-sequence together with this A-sequence in a shrinking generator, and the produced sequence will be non-periodic. Secondly, we show that there are what we call universal bi-ideals – finitely generated bi-ideals that generate non-periodic pseudo-random sequences when used as the S-sequence in a shrinking generator with any A-sequence containing both zeroes and ones. We give a description of a class of such universal bi-ideals.

Balkova et al. states that this modificated shrinking generator has not passed some latest statistical tests. They recently introduced a combinatorial condition called well distributed occurrences, or WELLDOC for short in (Balková et al., 2013a) and (Balková et al., 2013b) for that reason. In both papers they state that an infinite word with the property of well distributed occurrences (WELLDOC) is used to combine two linear congruential generators and form an infinite aperiodic sequence with good statistical behavior. The WELLDOC property is quite strong because it requires some forceful properties for every factor and all the integers. It demands all factors to be well distributed and since bi-ideals have a definite structure we tried to find some conditions for bi-ideals to satisfy well distributed occurrences. We proved that there exists a 1-bounded bi-ideal over the finite alphabet that satisfies the WELLDOC property. Furthermore, the given construction in Chapter 3 permits to construct infinitely many such 1-bounded bi-ideals with a such property.

Linearly recurrent bounded bi-ideals

We consider another interesting property of infinite words – linear recurrence in this thesis. An infinite word is linearly recurrent if it is uniformly recurrent and there exists a constant K such that the return time to an arbitrary its factor u is bounded by $K|u|$. In other words, the gap between two consecutive occurrences of a factor of length n does not exceed $K \cdot n$. The linear recurrence of the infinite word implies the linearity of its subword complexity (Durand et al., 1999). For morphic sequences (for a survey on morphic words see (Allouche and Shallit, 2003)) the uniform recurrence is equivalent to the linear recurrence (see (Durand, 1998) and (Durand et al., 2013)). As finitely generated bi-ideals are morphic words (for construction see, e.g., (Cers, 2012)), then finitely generated bi-ideals are linearly recurrent. We also give a characterization of linearly recurrent bounded bi-ideals. We introduce the notion of completely bounded bi-ideals and prove that completely bounded bi-ideals are exactly linearly recurrent bounded bi-ideals. This class is very large, namely, its cardinality is continuum.

Partial bi-ideals and finding of basis

In nowadays the information is what all is about and most of the time we do not have all the information we need. The desire and importance of getting back the lost information or revealing some unknown one is growing very fast. It is an axiom that all the information can be converted into words. In case we do not have some information, we get to so called partial words. It all started almost 15 years ago, when Blanchet-Sadri et al. (Blanchet-Sadri and Hegstrom, 2002) combined partial words with the well known theorem of Fine and Wilf 25. The work in the field of partial words has been tremendous from the side of Blanchet-Sadri. Her research on partial words has many edges, but most important are periodicity, for example (Blanchet-Sadri and Chriscoe, 2004) and (Blanchet-Sadri et al., 2008), complexity, for example (Blakeley et al., 2009) and (Blanchet-Sadri et al., 2012) and avoidabiliby, for example (Blanchet-Sadri et al., 2009) and (Blanchet-Sadri et al., 2012).

Blanchet-Sadri et al. in (Blanchet-Sadri and Hegstrom, 2002) accented that partial words appear in natural ways in several fields such as DNA computing, data communication, molecular biology etc. This was the inspiration of Chapter 5 – aggregate partial words with the class of infinite words, what we are interested in, i.e., bi-ideals. Nowadays the importance of information is so expansive that it is not possible to overvalue it and there are times and reasons of not knowing the full information about something, for example, DNA structure. As DNA have some kind of structure (with possibility of missing information) and bi-ideals (in this case, finitely generated bi-ideals) have a structure, in this chapter we are trying to solve the problem of filling the holes (missing information) in finitely generated bi-ideals. In general case of finite amount of holes in finitely generated bi-ideals it is always possible to get all the information back. We prove that in general case of infinite amount of holes it is not possible.

(Cers, 2012) solved the decision problem: given two basis, decide whether they generate the same finitely generated bi-ideal. In this case the given part was basis, so we turned the problem around and supposed that the given part is the bi-ideal. We were interested in a way opposite problem – can we found a basis for a given finitely generated bi-ideal? As it turned out, for a given finitely generated bi-ideal it is always possible.

The introduction to a new metric on the set of infinite words

We give an insight for measure and metrics in combinatorics on words. A good insight about topologies on words gives (Calude et al., 2009). Mostly in combinatorics on words two types of metrics are used (see, e.g., (Allouche and Shallit, 2003) and (Holmgren, 2000)). These types of metrics give poor information and have some shortages, shown by the help of some examples. A new approach of metric (fuzzy metric) in combinatorics on words has been introduced in Chapter 6. We try to justify the advantages of the use of fuzzy metrics instead of the ordinary metric for the description of the nearness-type structures on the set of infinite words.

Goals and objectives

The main objectives of this thesis is to research the essential properties and applications of finitely generated bi-ideals and bounded bi-ideals and to find a new approach for measuring infinite words.

The tasks of the thesis therefore are:

- to explore possible applications of finitely generated bi-ideals and bounded bi-ideals in cryptography;
- to construct an algorithm or a procedure that creates bounded bi-ideals with WELLDOC property;
- to solve the problem of filling of the holes in a finitely generated bi-ideal;
- to construct and describe a new metric on the set of infinite words.

The scientific importance of the thesis

In the thesis I offer a procedure that allows to create infinitely many bounded bi-ideals that have WELLDOC property. I describe an algorithm for finding a basis for a given finitely generated bi-ideal. The problem of filling of holes (”do not know” symbols) in finitely generated bi-ideal is solved. A new metric on the set of infinite words has been introduced. It has been shown that such (fuzzy) metric gives better description for closeness of infinite words.

Together with co-authors we give an approach of a non-periodic random number generator. We introduce the notion of a completely bounded bi-ideal and we prove that a bounded bi-ideal is linearly recurrent if and only if it is completely bounded.

The structure of the thesis

- Chapter 1 gives basic definitions and creates the background for the whole thesis. It also contains some already known results on finitely generated bi-ideals and bounded bi-ideals and some of them are used later in the thesis.
- In Chapter 2 we give a modification of the shrinking generator. With that we obtain aperiodic pseudo-random sequences.
- Chapter 3 is a natural continuation of chapter 2. It introduces the WELLDOC property, which is more advanced and passes more statistical tests than the shrinking generator. The existence of a 1-bounded bi-ideal that satisfies the WELLDOC property has been proved in this chapter.
- Chapter 4 gives the notion of completely bounded bi-ideals and its connection with linear recurrence. In this chapter we prove that a bounded bi-ideal is linearly recurrent if and only if it is completely bounded.
- Chapter 5 introduces so called partial words, which have many potentional applications in other scientifical fields. Possibility of finding the basis for a given finitely generated bi-ideal sequence in this chapter is given. Also possibility to fill the finite number of holes for a given finitely generated bi-ideal is proved. Divergence of two irreducible finitely generated bi-ideals has been explored in this chapter.
- Chapter 6 introduces measure and metric in combinatorics on words. A new approach of metric (fuzzy metric) in combinatorics on words has been introduced in this chapter.

Advantages of fuzzy metric have been shown.

Approbation

The results obtained during the thesis writing process have been presented at 4 international conferences and 4 domestic conferences (see full list on page 88). Results on possible use of finitely generated bi-ideals in cryptography have been presented at the 13th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing in Timisoara, Romania (2011). In Romania, the results were presented by I.Bērziņa. Results on the measure of some classes of infinite words have been presented at the 14th Mons Days of Theoretical Computer Science in Louvain-La-Neuve, Belgium (2012). Results on the relation of bounded bi-ideals and linearly recurrent words have been presented at the 15th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing in Timisoara, Romania (2013). Results on the existence of a 1-bounded bi-ideal that satisfies the WELLDOC property have been presented at 15th Central European Conference on Cryptology in Klagenfurt, Austria (2015).

A list of author's publications is given at the end of the bibliography.

1 Preliminaries and Background

1.1 Preliminaries

Let $\mathbb{N}$ denote the set of all non-negative integers. Let $\mathbb{N}_+ = \mathbb{N} \setminus \{0\}$. By Σ_n we denote the set $\{0, 1, \ldots, n-1\}$ for some $n \in \mathbb{N}_+$. Let $\overline{i, j}$ be the set $\{i, i+1, \ldots, j-1, j\}$, where i and j are two non-negative integers such that $i \leq j$.

Let A be a finite non-empty set called an *alphabet*. The elements of A are called *letters*. A string of letters $u = a_0 a_1 \cdots a_{n-1}$ from A is called a *finite word* of *length* n. We denote the length of a finite word u by $|u|$ and the number of occurrences of a letter $a \in A$ in a word u by $|u|_a$. We denote the empty word by λ and define $|\lambda| = 0$. By A^* and A^+ we denote the sets of all finite words and all finite non-empty words over alphabet A, respectively. For finite words $u = a_0 a_1 \cdots a_n$ and $v = b_0 b_1 \cdots b_m$ we say that a word $uv = a_0 a_1 \cdots a_n b_0 b_1 \cdots b_m$ is the *concatenation of* u *and* v. A word w' is called a *factor* of $w \in A^*$ if there exist $u, v \in A^*$ such that $w = uw'v$. The word u (v, respectively) is called a *prefix* (*suffix*, respectively) of w. By u^n we denote the finite word $\underbrace{uu \cdots u}_{n}$, where $n \in \mathbb{N}_+$ and u is a finite word.

A total map $x : \mathbb{N} \to A$ is called a *(right) infinite word* and the set of all infinite words is denoted by A^ω. For all $i \geq 0$ we set $x_i = x(i)$ and write simply

$$x = x_0 x_1 \cdots x_n \cdots$$

The notion of suffix, prefix and factor generalizes straightforwardly to infinite words by setting that suffix v is an infinite word. Also concatenation extends naturally to the case when the right word is infinite. By $x[i, j]$ we denote a factor of a word x starting in the position i and ending in the position j, where $i < j$, in other words, $x[i, j] = x_i x_{i+1} \cdots x_{j-1} x_j$. We denote $x[i, j) = x_i x_{i+1} \cdots x_{j-1}$, where $i + 1 < j$. A non-negative integer i is called an *occurence* of a word u in a word x if $x[i; i + |u|) = x[i; i + |u| - 1] = u$. If u is a factor of x, then we also say that u occurs or appears in x. We also write $u \searrow x$ if u appears in x.

An infinite word $x = x_0x_1 \dots x_n \dots$ is called *periodic* with a *period* p if $x_i = x_{i+p}$ for all $i \in \mathbb{N}$. If x is periodic with a period p and $v = x_0x_1 \cdots x_{p-1}$, where $x_i \in A$ for all $i \in \mathbb{N}$, we write $x = v^\omega$. A word is called *non-periodic* if it is not periodic. A word x is called *ultimately periodic* if there exist words $u \in A^*$, $v \in A^+$ such that $x = uv^\omega$. Each periodic word $x = v^\omega$ is ultimately periodic, since it can be written in the form $x = uv^\omega$, where $u = \lambda$. A word is called *aperiodic* if it is not ultimately periodic.

An infinite word is called *recurrent* if each of its factors occurs in it infinite number of times. An infinite word x is called *uniformly recurrent* if for each non-negative integer n there exists a non-negative integer m such that each x factor of length m contains as factors all factors of x of length n.

A sequence of finite words $v_0, v_1, \dots, v_i, \dots$ is called a *bi-ideal sequence* if for each $i \geq 0$, $v_{i+1} \in v_iA^*v_i$ and $v_0 \neq \lambda$. If $v_0, v_1, \dots, v_n, \dots$ is a bi-ideal sequence, then there exists a unique sequence of finite words $u_0, u_1, \dots, u_n, \dots$ with $u_0 \neq \lambda$ called the *basis* of the bi-ideal sequence (v_n) such that

$$v_0 = u_0$$

$$v_{i+1} = v_iu_{i+1}v_i.$$

The infinite word one gets as a limit of this bi-ideal sequence $x = \lim_{n\to\infty} v_n$ is called a *bi-ideal* and the sequence (u_n) is called a the basis of x or, equivalently, we say that (u_n) generates x. We also say that (u_n) generates the bi-ideal sequence (v_n).

Definition 1. The bi-ideal is called *finitely generated* if its basis sequence (u_i) is periodic, i.e., there exists a positive integer m such that for all $i, j \in \mathbb{N}$,

$$i \equiv j \pmod{m} \Rightarrow u_i = u_j.$$

In this case we say that the $m-$tuple $\langle u_0, u_1, \dots, u_{m-1}\rangle$ is a *finite basis* (or just a basis for short) of the finitely generated bi-ideal x. We also say that the basis $\langle u_0, u_1, \dots, u_{m-1}\rangle$ generates the bi-ideal sequence (v_n).

Example 1. If the $3-$tuple $\langle 0, 1, 1\rangle$ is a basis of the bi-ideal x, then

$$\begin{aligned} v_0 &= u_0 = 0, \\ v_1 &= v_0u_1v_0 = 010, \\ v_2 &= v_1u_0v_1 = 0101010, \\ v_3 &= v_1u_0v_1 = 010101000101010, \\ &\dots \\ x &= 0101010001010101010101000101010\cdots . \end{aligned}$$

Definition 2. If $(u_n)_{n\geq 0}$ is the basis of a bi-ideal x and there exists a non-negative integer l such that, for each i, $|u_i| \leq l$, then the bi-ideal x is called *bounded*.

1.2 Some known results on bi-ideals

In this section some known results on bi-ideals are given.

Proposition 1. *An infinite word x is recurrent if and only if it is a bi-ideal.*

Lemma 2. *Let $x \in A^\omega$ be an ultimately periodic word. If x is recurrent, then it is periodic.*

Due to the Proposition 1 and Lemma 2 in case of bi-ideals terms non-periodicity and aperiodicity are equivalent.

Proposition 1 and Lemma 2 gives that terms non-periodicity and aperiodicity are equivalent in bi-ideal case.

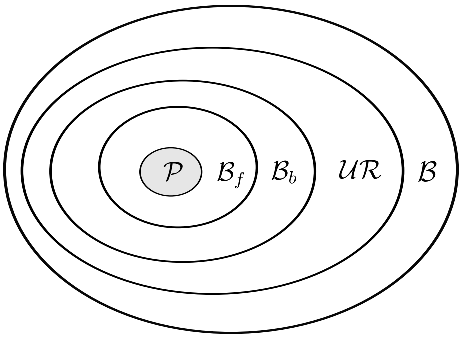

Figure 1.1: Hierarchy of the class of bi-ideals

(Buls and Lorencs, 2008) considered the hierarchy (see Figure 1.1):

$$\mathcal{P} \subset \mathcal{B}_f \subset \mathcal{B}_b \subset \mathcal{UR} \subset \mathcal{B},$$

where

$\mathcal{P}$ – the class of periodic words,

$\mathcal{B}_f$ – the class of finitely generated bi-ideals,

$\mathcal{B}_b$ – the class of bounded bi-ideals,

$\mathcal{UR}$ – the class of uniformly recurrent words,

$\mathcal{B}$ – the class of bi-ideals.

Theorem 3. *A bi-ideal x is periodic if and only if*

$$\exists n \in \mathbb{N} \exists u \exists v \left(v_n u \in v^* \wedge \forall i \in \mathbb{N}_+ u_{n+i} \in uv^* \right).$$

Theorem 4. *Let (u_i) be a sequence of words, which contains every u_j infinitely often. The bi-ideal x generated by (u_i) is periodic if and only if*

$$\exists w \forall i \left(u_i \in w^* \right).$$

Theorem 5. *A bi-ideal $x \in A^\omega$ that is generated by an m-tuple $\langle u_0, u_1, ..., u_{m-1} \rangle$ is periodic if and only if there exists a finite word $w \in A^+$ such that for all $i \in \overline{0, m-1}$*

$$u_i \in w^*.$$

(Lorencs, 2012) showed how to change the basis sequence of a finitely generated bi-ideal and proved that each finitely generated bi-ideal has countably many bases with the same number of basis words. In fact, his construction can be used for changing the basis sequence of an arbitrary bi-ideal.

Proposition 6. *If x is a bi-ideal generated by a sequence $(u_n)_{n \geq 0}$, then the sequence $u'_0, u'_1, \dots, u'_n, \dots$, where $u'_i = u_0 u_{i+1}$, also generates x.*

Later in Chapter 5 we will show a bit different change of basis for finitely generated bi-ideals, which contains getting rid of the first letter.

Example 2. If x is a bounded bi-ideal with a basis sequence $0, 1, 00, 00, 00, \dots$, then sequences $01, 000, 000, 000, 000, \dots$ and $01000, 01000, 01000, 01000, 01000, \dots$ are also basis sequences of x.

Proposition 7. *If $\langle u_0, u_1, \dots, u_{m-1} \rangle$ is a basis of a finitely generated bi-ideal x, then the m-tuple $\langle u'_0, u'_1, \dots, u'_{m-1} \rangle$, where $u'_i = u_0 u_s$ and $s = i + 1 \bmod m$, also is a basis of x.*

Corollary 8. *Every finitely generated bi-ideal x has countably many bases with the same number of basis words.*

Example 3. Let $\langle 0, 1, 2\rangle$ be a basis of a finitely generated bi-ideal x. Then 3-tuples $\langle 01, 02, 00\rangle$, $\langle 0102, 0100, 0101\rangle$, and $\langle 01020100, 01020101, 01020102\rangle$ are also bases of x.

If Proposition 6 or Corollary 7 is applied to some basis sequence of a bi-ideal x, then we say that basis words of the bi-ideal x are L-prolonged or simply that the basis sequence of the bi-ideal x (or the basis of a finitely generated bi-ideal) is L-prolonged. If x is a bi-ideal with a basis sequence $u_0, u_1, \ldots, u_n, \ldots$, then for all $n > 0$ the sequence

$$u_0^{(n)}, u_1^{(n)}, \ldots, u_m^{(n)}, \ldots,$$

where $u_i^{(n)} = u_0^{(n-1)} u_{i+1}^{(n-1)}$, is *the basis sequence of the bi-ideal x after n iterations of L-prolongation*. If x is a finitely generated bi-ideal with basis $\langle u_0, u_1, \ldots, u_{m-1}\rangle$, then for all $n > 0$ the m-tuple

$$\left\langle u_0^{(n)}, u_1^{(n)}, \ldots, u_{m-1}^{(n)} \right\rangle,$$

where $u_i^{(n)} = u_0^{(n-1)} u_{i+1 \bmod m}^{(n-1)}$, is *the basis of the finitely generated bi-ideal x after n iterations of L-prolongation*.

Lemma 9. *Let $(u_n)_{n\geq 0}$ be a basis sequence of a bi-ideal x. Let $(v_n)_{n\geq 0}$ be the bi-ideal sequence generated by $(u_n)_{n\geq 0}$. Then for each $n \in \mathbb{N}_+$ and each $i \in \mathbb{N}$*

$$u_i^{(n)} = v_{n-1} u_{i+n}.$$

Proof. If $n = 1$ then, by definition of L-prolongation, $u_i^{(1)} = u_0 u_{i+1} = v_0 u_{i+1}$ for each $i \in \mathbb{N}$. Assume that for all $n \leq k$ and for all $i \in \mathbb{N}$ we have $u_i^{(n)} = v_{n-1} u_{i+n}$. Let us prove that it also holds for $n = k + 1$. By definition of L-prolongation and our assumption, for all $i \in \mathbb{N}$ we have

$$u_i^{(k+1)} = u_0^{(k)} u_{i+1}^{(k)} = v_{k-1} u_{0+k} v_{k-1} u_{i+1+k} = v_k u_{i+k+1}.$$

□

Corollary 10. *Let x be a bounded bi-ideal. Let $(u_n)_{n\geq 0}$ be a basis sequence of x such that each element of $(u_n)_{n\geq 0}$ occurs in $(u_n)_{n\geq 0}$ infinite number of times. Then for all $k \geq 1$ each element of the basis sequence $(u_n^{(k)})_{n\geq 0}$ occurs in $(u_n^{(k)})_{n\geq 0}$ infinitely often.*

Proof. It follows from Lemma 9. □

Lemma 11. *If x is a bounded bi-ideal, then there exists a basis sequence $(u_n)_{n\geq 0}$ of x such that each element of $(u_n)_{n\geq 0}$ occurs in $(u_n)_{n\geq 0}$ infinite number of times.*

Proof. Let x be a bounded bi-ideal with a basis sequence $(u_n)_{n\geq 0}$. Since the length of each basis word is bounded by some $\ell \in \mathbb{N}_+$, then there is at least one basis word u_i that occurs in $(u_n)_{n\geq 0}$ infinitely many times. Hence there is a non-negative integer δ such that each element of the sequence $(u_{\delta+k})_{k\geq 0}$ occurs in $(u_{\delta+k})_{k\geq 0}$ infinite number of times.

We L-prolong the basis δ times. Then by Lemma 9 we have

$$u_k^{(\delta)} = v_{\delta-1}u_{k+\delta}$$

for all $k \in \mathbb{N}$. Since $v_{\delta-1}$ is a common prefix of words $u_0^{(\delta)}, u_1^{(\delta)}, \ldots, u_n^{(\delta)}, \ldots$ and since for each $k \in \mathbb{N}$ the basis word $u_{k+\delta}$ occurs in the sequence $(u_{\delta+n})_{n\geq 0}$ infinitely many times, the basis sequence $(u_n^{(\delta)})_{n\geq 0}$ satisfies conditions of the lemma. □

2 On a Non-periodic Shrinking Generator

2.1 Preliminaries

We start with a well–known approach (see, e.g., (L'Ecuyer, 1998)). As of today, the most convenient and reliable way of generating the random numbers for stochastic simulations appears to be via deterministic algorithms with a solid mathematical basis. These algorithms produce sequences of bits which are, in fact, not random at all, but seem to behave as if the bits were chosen independently at random.

Definition 3. A pseudo-random number generator is a structure $\mathfrak{G} = \langle Q, B, q_0, T, G \rangle$, where Q is a finite set of states, $q_0 \in Q$ is the initial state (or seed), the mapping $Q \xrightarrow{T} Q$ is the transition function, B is finite set of symbols, and $Q \xrightarrow{G} B$ is the output function.

This model is called a Moore machine in automata theory. In fact, this model is the specialised Moore machine. The state of a generator is initially q_0 and evolves according to the recurrence $q_n = T(q_{n-1})$, for $n = 1, 2, 3, ...$. At step n the generator outputs the symbol $b_n = G(q_n)$.

Clearly, since the state space Q is finite, the sequence of states q_n is ultimately periodic; therefore, this approach is limited. One method for obtaining non-periodic sequences is to use the simplest chaotic system — the logistic map. In 1982 Oishi and Inoue (Oishi and Inoue, 1982) proposed the idea to use chaos in designing a pseudo-random generator. In 1992 Sandri introduced a simple non-periodic pseudo-random number generator which is based on a simple logistic map (see (Sandri, 1992)). Recently, Hu et. al. (Hu et al., 2009) proposed a true random number generator by combining congruential methods with prime numbers and higher order composition of logistic maps. It generates a 256-bit random number by computer mouse movement. For more information of using chaotic systems in generation of pseudo-random sequences, see e.g. (Patidar et al., 2009), (Phatak and Suresh Rao, 1995).

A pseudo-random number generator can be created by substituting the S-sequence by a finitely generated bi-ideal — a non-periodic sequence (see (Buls and Lorencs, 2008)). Obviously, such model is a generalization of the pseudo-random number generator (see definition 3). We conjecture, that for most non-trivial cases the resulting pseudo-random sequence is non-periodic. The resulting pseudo-random sequence has good statistical properties as indicated by the Diehard test suite (see Section 2.2.2).

In these theses we show two approaches for generating non-periodic pseudo-random number sequences using our modified shrinking generator. Firstly, given a periodic A-sequence, we prove that any finitely generated bi-ideal that satisfies a simple condition can be used as the S-sequence together with this A-sequence in a shrinking generator, and the produced sequence will be non-periodic. Secondly, we show that there are what we call universal bi-ideals — finitely generated bi-ideals that generate non-periodic pseudo-random sequences when used as the S-sequence in a shrinking generator with any A-sequence containing both zeroes and ones. We give a description of a class of such universal bi-ideals.

Definition 4. Let $x, y \in \{0,1\}^\omega$ be two infinite words with $|y|_1 = \infty$. The *shrunk sequence* of x by y is defined inductively:

$$w_1 := \begin{cases} x_1, & \text{if } y_1 = 1, \\ \lambda, & \text{if } y_1 = 0, \end{cases},$$

$$w_i := \begin{cases} w_{i-1}x_i, & \text{if } y_i = 1, \\ w_{i-1}, & \text{if } y_i = 0, \end{cases}$$

The infinite word $z = \lim_{i\to\infty} w_i$ is called the *shrunk word* of x by y and denoted by $z := S_y(x)$.

By $alph(u)$ we denote the set of distinct letters in the word u, i.e., $alph(u) = \{a \mid a \in A \wedge a \in \mathrm{F}(u)\}$. If x is an infinite non-empty word and $|alph(x)| = 1$, then x is called a *trivial word*, otherwise x is called a *non-trivial word.* Further we only consider non-trivial infinite words.

2.2 Non-periodic shrunk words

In this section we show a method for the construction of an infinite number of finitely generated bi-ideals from a given A-sequence, such that the corresponding shrunk sequence using the bi-ideal as the S-sequence is non-periodic. Afterwards, we shortly analyse test results.

2.2.1 Construction

In order to construct a non-periodic shrunk sequence, the finitely generated bi-ideal, which is used as S-sequence, has to be non-periodic. In 2008 Buls and Lorencs (Buls and Lorencs, 2008) obtained sufficient conditions for a finitely generated bi-ideal to be non-periodic:

Theorem 12. *If $\bigcup_{i=0}^{m-1}$ Pref(u_i) or $\bigcup_{i=0}^{m-1}$ Suff(u_i) has at least two words with the same length, then the bi-ideal with basis $\langle u_0, u_1, ..., u_{m-1}\rangle$ is non-periodic.*

However, the non-periodicity of the bi-ideal (S-sequence) alone is not a sufficient condition for the shrunk sequence to be non-periodic. Next, we give two examples (without proof), where the resulting sequence is periodic.

Example 4. If $x = (1100)^\omega$ and y is the finitely generated bi-ideal with basis $\langle 01, 10\rangle$, then $z = S_y(x) = (10)^\omega$.

Example 5. If $x' = (01)^\omega$ and y' is a finitely generated bi-ideal with basis $\langle 101, 10001\rangle$, then $z' = S_{y'}(x') = (0011)^\omega$.

In both examples Theorem 12 is satisified, e.g., the bi-ideals used as the S-sequences are non-periodic, but the resulting shrunk sequence is periodic. Moreover, the period of the shrunk sequence can be smaller or larger than the period of the respective A-sequence.

In order to construct a non-periodic shrunk sequence, we have to put some additional restrictions on the basis of the finitely generated bi-ideal that will be used as the S-sequence. First, we state two lemmata that will be used in the proof of main result of this section.

Lemma 13. *If $x \in \{0,1\}^\omega$ is a bi-ideal generated by $\langle u_0, u_1, ..., u_{m-1}\rangle$, then $\forall p, T \in \mathbb{N} \overset{\infty}{\exists} \alpha, \beta \in \mathbb{N}, \alpha \neq \beta$:*

$$|v_{\alpha m-1}| \equiv |v_{\beta m-1}| \textit{ (mod p)}, \tag{2.21}$$

$$|v_{\alpha m-1}|_1 \equiv |v_{\beta m-1}|_1 \textit{ (mod T)}, \tag{2.22}$$

where v_i denotes the i-th element of the bi-ideal sequence with the basis (u_n).

Proof. Let (v_n) be the bi-ideal sequence corresponding to the finitely generated bi-ideal x. We consider the subsequence $(v_{im-1})_{i\geq 1}$ of (v_n). Since (v_n) is an infinite sequence, $(v_{im-1})_{i\geq 1}$ is also an infinite sequence.

We partition $(v_{im-1})_{i\geq 1}$ into equivalence classes by their length modulus p:

$$\forall k \geq 1 \;\; A_t = \big\{v_{km-1} \big| \; |v_{km-1}| \equiv t \text{ (mod } p)\big\}. \tag{2.23}$$

Since $(v_{im-1})_{i\geq 1}$ is an infinite sequence, there exists an integer $\ell \in \{0, 1, \ldots, p-1\}$ such that $|A_\ell| = \infty$. For all $v_{k_1 m-1}, v_{k_2 m-1} \in A_\ell$ condition (2.21) holds.

Next, we partition $(v_{im-1})_{i\geq 1}$ further based on the number of ones modulo T:

$$\forall k \geq 1 \;\; B_t = \{v_{km-1} |\, v_{km-1} \in A_\ell \wedge |v_{km-1}|_1 \equiv t \pmod{T}\}.$$

Since $|A_\ell| = \infty$, there exists an integer $s \in \{0, 1, \ldots, T-1\}$, such that $|B_s| = \infty$. For all $v_{k_1 m-1}, v_{k_2 m-1} \in B_s$ conditions (2.21) and (2.22) hold. □

Lemma 14. *(Let (v_n) be a bi-ideal sequence, then*

$$\forall m \leq n \; v_m \in \mathit{Pref}(v_n) \cap \mathit{Suff}(v_n).$$

Now we state the main results of this section.

Proposition 15. *If x is a non-trivial infinite periodic word, then there exists an infinite number of finitely generated bi-ideals y, such that $z = S_y(x)$ is aperiodic.*

Proof. Let $x = u^\omega \in \{0,1\}^\omega$, where $|u| = p$. Let $y \in \{0,1\}^\omega$ be an aperiodic bi-ideal generated by $\langle u_0, u_1, \ldots, u_{m-1}\rangle$.
We will show a condition on the basis of y, such that the shrunk word $z = S_y(x)$ is aperiodic.

Suppose the contrary that the shrunk sequence is ultimately periodic, e.g., $z = v'v^\omega$ (where $|v'| = T_1$ and $|v| = T$). Then by lemma 13 we can choose $\alpha, \beta \in \mathbb{N}$ ($\alpha < \beta$) such that

$$|v_{\alpha m-1}| \equiv |v_{\beta m-1}| \quad \bmod p,$$
$$|v_{\alpha m-1}|_1 \equiv |v_{\beta m-1}|_1 \quad \bmod T,$$
$$|v_{\alpha m-1}| \geq p \wedge |v_{\alpha m-1}|_1 \geq T \wedge |v_{\alpha m-1}|_1 > T_1.$$

Therefore, there exist $k, k_1 \in \mathbb{N}$, such that both

$$|v_{\beta m-1}| - |v_{\alpha m-1}| = kp \tag{2.24}$$

$$|v_{\beta m-1}|_1 - |v_{\alpha m-1}|_1 = k_1 T \tag{2.25}$$

hold. Now we observe that $v_{\alpha m} = v_{\alpha m-1}u_0 v_{\alpha m-1}$ and $v_{\beta m} = v_{\beta m-1}u_0 v_{\beta m-1}$. Therefore from

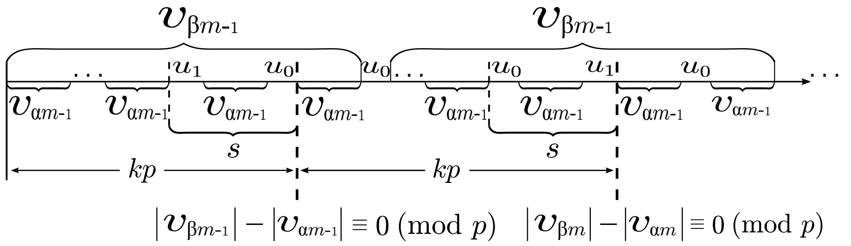

Figure 2.1: Structure of the bi-ideal y.

(2.24) and (2.25) and using lemma 14 we obtain (see Figure 2.1)

$$\begin{aligned}
&|y[|v_{\beta m-1}| - |v_{\alpha m-1}| + 1, |v_{\beta m}| - |v_{\alpha m}|]| = \\
&= |v_{\beta m}| - |v_{\alpha m}| - |v_{\beta m-1}| + |v_{\alpha m-1}| = \\
&= 2|v_{\beta m-1}| + |u_0| - 2|v_{\alpha m-1}| - \\
&-|u_0| - |v_{\beta m-1}| + |v_{\alpha m-1}| = \\
&= |v_{\beta m-1}| - |v_{\alpha m-1}| = kp
\end{aligned} \tag{2.26}$$

and

$$|y[|v_{\beta m-1}| - |v_{\alpha m-1}| + 1, |v_{\beta m}| - |v_{\alpha m}|]|_1 = |v_{\beta m-1}|_1 - |v_{\alpha m-1}|_1 = k_1 T. \tag{2.27}$$

Now, if we set $x' = x[1, kp]$, $y' = y[1, kp]$, $x'' = x[kp+1, 2kp]$, $y'' = y[kp+1, 2kp]$ and consider the shrinking construction for these finite fragments, then by (2.26) and (2.27) we obtain

$$S_{y'}(x') = v'z[T_1 + 1, k_1 T], \tag{2.28}$$

$$S_{y''}(x'') = v''z[k_1 T + T_1 + 1, 2k_1 T], \tag{2.29}$$

where $|v'| = |v''| = T_1$. Next, by (2.28), (2.29), $|v_{\alpha m-1}|_1 > T_1$ and from the assumption that z is ultimately periodic it follows that

$$z[T_1 + 1, k_1 T] = z[k_1 T + T_1 + 1, 2k_1 T], \tag{2.210}$$

Similarly, since $|u_1 v_{\alpha m-1} u_0| = |u_0 v_{\alpha m-1} u_1|$ and $|u_1 v_{\alpha m-1} u_0|_1 = |u_0 v_{\alpha m-1} u_1|_1$, it follows that

$$S_{y[kp-s+1,kp]}\big(x[kp - s + 1, kp]\big) = S_{y[2kp-s+1,2kp]}\big(x[2kp - s + 1, 2kp]\big), \tag{2.211}$$

where $s = |u_1 v_{\alpha m-1} u_0|$ (see Figure 2.2).

We will show how to construct u_0 and u_1 such that (2.211) does not hold, hence proving the existence of a finitely generated bi-ideal y, such that the shrunk word $z = S_y(x)$ is aperiodic.

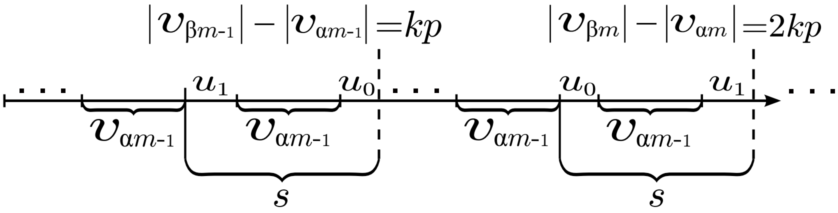

Figure 2.2: Structure of the bi-ideal y.

Since $|alph(x)| = 2$, it follows that

$$\exists i \in \overline{2,p} : (u[i-1] = a \wedge \forall j \in \overline{i,p}\, u[j] = \bar{a})), \tag{2.212}$$

where $a \in \{0,1\}$ and $\bar{a} = 1$ if $a = 0$ or $\bar{a} = 0$ if $a = 1$. We set

$$u_0 = u'10w,\ u_1 = u''01w, \tag{2.213}$$

where $w \in \{0,1\}^*$, $|w| = p - i$ and $u', u'' \in \{0,1\}^*$ are arbitrary finite words over the alphabet $\{0,1\}$.

If $|w|_1 = \gamma$, then by (2.212) and (2.213) $z[k_1T - \gamma] = a$ but $z[2k_1T - \gamma] = \bar{a}$. Thus (2.210) and (2.211) do not hold. Hence z is aperiodic. Since $u', u'' \in \{0,1\}^*$ are arbitrary finite words over alphabet $\{0,1\}$, there exists an infinite number of u_0, u_1 such that the shrunk word z is aperiodic.

Moreover, we have not made any restrictions on other elements of the basis of y. Therefore, for all $m \geq 3$ the basis words u_j $(j \geq 3)$ can be chosen arbitrarily. □

Corollary 16. *If x is a non-trivial infinite periodic word, then there exist an infinite number of finitely generated bi-ideals y, such that $z = S_y(x)$ is non-periodic.*

Proof. Since each periodic word is also ultimately periodic, the proof follows directly from the proof of the Proposition 15. □

2.2.2 Statistics

One way of evaluating the fitness of a pseudo random generator for cryptographic applications is to check whether the produced bit-sequence appears random in the statistical sense, i.e., that it does not exemplify patterns that would be unexpected in a sequence of truly random and independent coin flips. The simplest of such tests is the frequency test, that checks if the number of ones is close to the number of zeroes. Many such tests can and have been constructed and several software packages for testing pseudo-random number generators are available. We used the well known Diehard battery of tests (Marsaglia, 1996) to asses the fitness of our generator. This test suite includes 18 main and several more additional tests, all of which a good generator is expected to pass.

While it was known that the shrinking generator has good statistical properties (Coppersmith et al., 1994), this did not necessitate that these properties would carry over to our construction. Still, we found that our shrinking generator passed *all* tests in the Diehard test suite. For the testing purposes a 32 bit LFSR was taken as the A-sequence and a bi-ideal with base words of lengths around 2KB (that were generated by cutting up another 32-bit LFSR) was used as the S-sequence. Additionally the first two base words were altered in a way so that (2.213) was satisfied, making the shrunk sequence non-periodic (the required changes are very small compared to a freely selected base). The number of base words was not limited, but the lengths of the tests were such, that around the first 20 base words were used while performing each test.

2.3 Universal Bi-ideals

In Section 2.2 we showed how it was possible to construct non-periodic S-sequences for each periodic A-sequence such that the resulting shrunk words were non-periodic. Even though for each A-sequence there exists an infinite number of S-sequences such that the shrunk word is aperiodic, the choice of the S-sequence depends on the choice of the A-sequence. In order to simplify the choice of the sequences, it would be more convenient to use non-periodic bi-ideals (as S-sequences) such that for each non-trivial A-sequence the resulting shrunk word would be non-periodic. In Proposition 19 we prove the existence of such bi-ideals.

Definition 5. A Bi-ideal y is called *universal* if for all non-trivial periodic $x = u^\omega$, the shrunk word $z = S_y(x)$ is aperiodic.

Before turning to our main proposition, we will prove two easy but crucial lemmata:

Lemma 17. *Let $a, b \in A$, $u \in A^*$ and $|aub| > T > 1$. If T is the least period of aub, then $au \neq ub$.*

Proof. If $u = \lambda$, then $aub = ab$. Since $T > 1$ then $a \neq b$. Therefore

$$au = a \neq b = ub.$$

The rest of the proof is by induction on the length of the word u. Since T is the period of aub, the period t of the word au has to be less than or equal to T, i.e., $t \leq T$.

(i) If $t = 1$, then $au = a^n$, where $n = |au|$. Since $T > 1$ is the period of the word aub, $b \neq a$. Therefore $au = a^n \neq ub$.

(ii) Let $u = vc$ and $t > 1$, i.e., $t > 1$ is the period of the word $au = avc$. By the induction assumption $av \neq vc$. From this

$$au = avc \neq vcb = ub.$$

□

Lemma 18. *Let* $m \in \mathbb{N}$*,* $m \geq 2$*. If* $u_0 = 1$*,* $u_1 = 10$*,* $m > 2 \Rightarrow \big(\forall i \in \{2, 3, \ldots, m-1\}\ (00 \notin F(u_i))\big)$*, then* $00 \notin F(x)$*, where* x *is the bi-ideal generated by the basis* $\langle u_0, u_1, ..., u_{m-1}\rangle$*.*

Proof. The proof is by induction. By (v_n) we denoe the bi-ideal sequence generated by the basis $\langle u_0, u_1, \ldots, u_{m-1}\rangle$. Since $v_0 = 1$ and $v_1 = 1101$, then $00 \notin \mathrm{F}(v_0)$ and $00 \notin \mathrm{F}(v_1)$ and we assume that $00 \notin \mathrm{F}(v_i)$ for all $i \leq k$.

Since $v_{k+1} = v_k u_j v_k$, where $j \equiv k + 1$ (mod m) and both $00 \notin \mathrm{F}(v_k)$ and $00 \notin \mathrm{F}(u_j)$, and $1 = v_0 \in$ Pref$(v_k) \cap$ Suff(v_k) (by lemma 14), then $00 \notin \mathrm{F}(v_{k+1})$. □

Proposition 19. *Let* $m \in \mathbb{N}$*,* $m \geq 2$*. If* $u_0 = 1$*,* $u_1 = 10$ *and* $00 \notin F(u_i)$ *for all* $i \in \{2, 3, \ldots, m-1\}$*, then the bi-ideal generated by the basis* $\langle u_0, u_1, \ldots, u_{m-1}\rangle$ *is a universal bi-ideal.*

Proof. Let y be the bi-ideal generated by the m-tuple $\langle u_0, u_1, ..., u_{m-1}\rangle$. Let's assume the contrary that y is not a universal bi-ideal. Then there exists a non-trivial periodic word $x = u^\omega$ with $|u| = p \geq 2$, such that $z = S_y(x)$ is a ultimately periodic word with period T and pre-period T_1, i.e., $z = wv^\omega$, where $|v| = T$ and $|w| = T_1$.

By lemma 13, we can choose sufficiently large $\alpha, \beta, \gamma, \delta \in \mathbb{N}$, such that $|v_{\alpha m-1}|_1 > T_1$ and

$$|v_{\alpha m-1}| \equiv |v_{\beta m-1}| \equiv |v_{\gamma m-1}| \equiv |v_{\delta m-1}| \mod p,$$

$$|v_{\alpha m-1}|_1 \equiv |v_{\beta m-1}|_1 \equiv |v_{\gamma m-1}|_1 \equiv |v_{\delta m-1}|_1 \mod T,$$

$$|v_{\delta m-1}| > |v_{\gamma m-1}| > |v_{\beta m-1}| > |v_{\alpha m-1}| > p,$$

$$|v_{\delta m-1}|_1 > |v_{\gamma m-1}|_1 > |v_{\beta m-1}|_1 > |v_{\alpha m-1}|_1 > T,$$

which implies

$$|v_{\beta m-1}| - |v_{\alpha m-1}| = kp, \tag{2.31}$$

$$|v_{\beta m-1}|_1 - |v_{\alpha m-1}|_1 = k_1 T. \tag{2.32}$$

for some $k, k_1 \in \mathbb{N}$.

Now, similarly to the proof of Proposition 15, we observe that $v_{\alpha m} = v_{\alpha m-1}1v_{\alpha m-1}$ and $v_{\beta m} = v_{\beta m-1}1v_{\beta m-1}$ and, therefore, from (2.31), (2.32) and using lemma 14 we obtain (see Figure 2.3)

$$|y[|v_{\beta m-1}| - |v_{\alpha m-1}| + 1, |v_{\beta m}| - |v_{\alpha m}|]| = |v_{\beta m-1}| - |v_{\alpha m-1}| = kp \tag{2.33}$$

and

$$|y[|v_{\beta m-1}| - |v_{\alpha m-1}| + 1, |v_{\beta m}| - |v_{\alpha m}|]|_1 = |v_{\beta m-1}|_1 - |v_{\alpha m-1}|_1 = k_1 T. \tag{2.34}$$

Now, from the periodicity of x and the equations (2.33) and (2.34) we obtain

$$x[kp - |v_{\alpha m-1}|, kp - 1] = x[2kp - |v_{\alpha m-1}|, 2kp - 1], \tag{2.35}$$

$$y[kp - |v_{\alpha m-1}|, kp - 1] = y[2kp - |v_{\alpha m-1}| - 1, 2kp - 2] = v_{\alpha m-1}, \tag{2.36}$$

and,

$$|y[kp - |v_{\alpha m-1}|, kp]|_1 = |y[2kp - |v_{\alpha m-1}| - 1, 2kp]|_1. \tag{2.37}$$

If we set $|v_{\alpha m-1}| = \ell$ and consider the same shrinking construction for finite words

$$x' = x[kp - \ell, kp - 1],$$
$$x'' = x[2kp - \ell - 1, 2kp - 2] = x[kp - \ell - 1, kp - 2],$$
$$y' = v_{\alpha m-1},$$

then from here, (2.35), (2.36) and (2.37) and using our assumption that z is ultimately periodic we obtain

$$S_{v_{\alpha m-1}}(x[kp - \ell, kp - 1]) = S_{v_{\alpha m-1}}(x[kp - \ell - 1, kp - 2]). \tag{2.38}$$

If we further set $x[kp - \ell - 1, kp - 1] = avb = v' = v'_1 v'_2 \dots v'_{\ell+1}$, then

$$S_{v_{\alpha m-1}}(av) = S_{v_{\alpha m-1}}(vb), \tag{2.39}$$

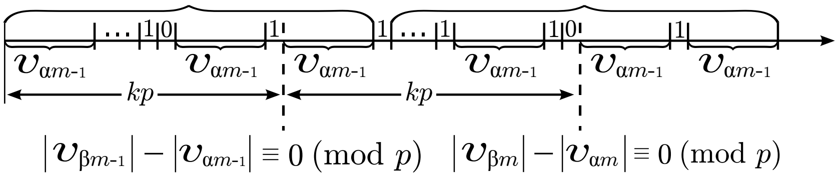

Figure 2.3: Structure of the bi-ideal y.

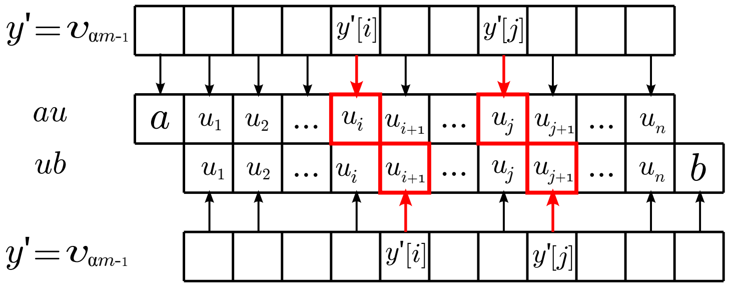

Figure 2.4: Structure of the bi-ideal y.

but $av \neq vb$ from lemma 17. From here

$$\exists i > 1 \forall j \leq i : \; v'[j-1] = v'[j] \wedge v'[i] \neq v'[i+1], \tag{2.310}$$

but from (2.39) it follows that

$$\forall s \in \overline{1, \ell} \; S_{y'[1,s]}(v'[1,s]) = S_{y'[1,s]}(v'[2,s+1]). \tag{2.311}$$

Observe that if i is the index mentioned in (2.310) and $y'[i] = 1$, then from (2.311) equation (2.39) does not hold (see Figure 2.4). Thus $y'[i] = 0$. Moreover, since (2.311) holds for all $s \in \{1, 2, \ldots, \ell\}$ then

$$\forall t \in \overline{1, \ell - 1} : \; v'[t] \neq v'[t+1] \Rightarrow y'[t] = 0. \tag{2.312}$$

Since $|alph(u)| = 2$ and $\ell > p$ there exists an index $t_0 < p$ such that (2.312) holds. From this, (2.312) and the periodicity of x we get that for all $t \in \{1, 2, \ldots, p-1\}$ and for all $\mu \in \mathbb{N}$

$$(v'[t] \neq v'[t+1] \wedge t + \mu p < \ell) \Rightarrow y'[t + \mu p] = y'[t] = 0, \tag{2.313}$$

i.e., there are zeros in $y' = v_{\alpha m-1}$ repeating periodically with period p.

Similarly, if we consider $v_{\gamma m-1}$ and $v_{\delta m-1}$ (instead of $v_{\alpha m-1}$ and $v_{\beta m-1}$), we obtain that there are zeros in $v_{\gamma m-1}$ that repeat periodically with period p. From this and considering $|alph(u)| = 2$ and $|v_{\gamma m-1}| > \ell > p$, there exists an index $i_0 < p$ such that for all $\eta \in \mathbb{N}$

$$i_0 + \eta p \leq |v_{\gamma m-1}| \Rightarrow v_{\gamma m-1}[i_0] = v_{\gamma m-1}[i_0 + \eta p] = 0. \tag{2.314}$$

From this and the fact that $v_{\alpha m-1} \in \text{Pref}\,(v_{\gamma m-1})$ it follows that for all $\eta \in \mathbb{N}$

$$i_0 + \eta p \leq |v_{\alpha m-1}| \Rightarrow v_{\alpha m-1}[i_0] = v_{\alpha m-1}[i_0 + \eta p] = 0. \tag{2.315}$$

Since $\alpha < \beta < \gamma$ and $m = 2$ then $|v_{\alpha m-1}| < |v_{\beta m-1}| < |v_{\beta m}| < |v_{\gamma m-1}|$. From this and the equations (2.31), (2.32) and (2.314) we obtain

$$v_{\gamma m-1}[i_0] = v_{\gamma m-1}[i_0 + kp] = v_{\gamma m-1}[i_0 + 2kp] = 0 \tag{2.316}$$

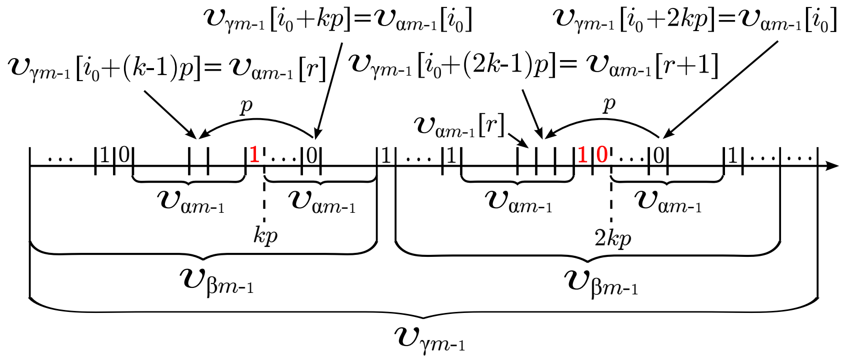
$v_{\gamma m-1}[i_0+kp]=v_{\alpha m-1}[i_0]$
$v_{\gamma m-1}[i_0+2kp]=v_{\alpha m-1}[i_0]$
$v_{\gamma m-1}[i_0+(k-1)p]= v_{\alpha m-1}[r]$
$v_{\gamma m-1}[i_0+(2k-1)p]= v_{\alpha m-1}[r+1]$
p
$v_{\alpha m-1}[r]$
p
… 1 0 1 … 0 1 … 1 1 0 … 0 1 … …
$v_{\alpha m-1}$
$v_{\alpha m-1}$
kp
$v_{\alpha m-1}$
$v_{\alpha m-1}$
$2kp$
$v_{\beta m-1}$
$v_{\beta m-1}$
$v_{\gamma m-1}$

Figure 2.5: Structure of the bi-ideal with basis $\langle 1, 10\rangle$.

and

$$v_{\gamma m-1}[i_0] = v_{\gamma m-1}[i_0 + (k-1)p] = v_{\gamma m-1}[i_0 + (2k-1)p] = 0. \tag{2.317}$$

Next we observe that from the construction of a bi-ideal (see Figure 2.1 and 2.5) and from the equations (2.31) and (2.32) it follows that

$$v_{\gamma m-1}[kp - \ell - 1, kp] = y[kp - \ell - 1, kp] = v_{\alpha m-1}1 \tag{2.318}$$

and

$$v_{\gamma m-1}[2kp - \ell - 2, kp] = y[2kp - \ell - 2, kp] = v_{\alpha m-1}10. \tag{2.319}$$

Since $v_{\gamma m-1}[kp-1] = v_{\alpha m-1}[\ell] = 1$ (by construcion $v_{\gamma m-1}[kp-1] = 1$ and by lemma 14 — $v_0 \in \mathrm{Suff}\,(v_{\alpha m-1})$), then $i_0 \neq p-1$ and $i_0 \neq p$.

Further, if $v_{\gamma m-1}[i_0 + (k-1)p] = v_{\alpha m-1}[r]$ (where $r \in \{1, 2, \ldots, \ell-1\}$), then $v_{\gamma m-1}[i_0 + (2k-1)p] = v_{\alpha m-1}[r+1]$. Finally from (2.317) it follows that $v_{\alpha m-1}[r] = v_{\alpha m-1}[r+1] = 0$, i.e., $00 \in \mathrm{F}(y)$, but from lemma 18 we know that $00 \notin \mathrm{F}(y)$. This is a contradiction and therefore $z = S_y(x)$ is not ultimately periodic. □

3 WELLDOC property in bi-ideals

3.1 Preliminaries

Balkova et al. states that the modified shrinking generator, introduced in Chapter 2, has not passed some latest statistical tests. They recently introduced a combinatorial condition called well distributed occurrences, or WELLDOC for short in (Balková et al., 2013a) and (Balková et al., 2013b) for that reason. In both papers they state that an infinite word with the WELLDOC property is used to combine two linear congruential generators and form an infinite aperiodic sequence with good statistical behavior (better that the modificated shrinking generator).

This was the main reason why the author has analysed the WELLDOC property for bounded bi-ideals, a subclass of recurrent words, and has proved the existence of a 1-bounded bi-ideal over the finite alphabet that satisfies the WELLDOC property in this thesis (particularly, this chapter). Significantly that the procedure given in this chapter allows to construct infinitely many such 1-bounded bi-ideals with such a property.

Bounded bi-ideals are natural extension of finitely generated bi-ideals. Let $S \subseteq A^*$ be a finite set of words from the alphabet A. Pick up randomly or algorithmically a sequence of words (u_i), where $u_i \in S$ for all $i \in \mathbb{N}$. This sequence generates a bounded bi-ideal $x = \lim_{n\to+\infty} v_n$ with $v_0 = u_0$ and $v_n = v_{n-1}u_n v_{n-1}$ for all n greater than 0. Clearly, if (u_i) is periodic, then we obtain a finitely generated bi-ideal.

For a finite or infinite word $u = u_0u_1u_2\ldots$, $Pref_n u$ will denote the prefix of length n of u, i.e., $Pref_n u = u_0u_1\ldots u_{n-1}$. The *Parikh vector* of a finite word over an alphabet $\{0, 1, \ldots, d-1\}$ is defined as $(|w|_0, |w|_1, \ldots, |w|_{d-1})$.

Let $i_0, i_1, \ldots$ denote the occurrences of w in an aperiodic infinite word u over the alphabet $\{0, 1, \ldots, d-1\}$. According to the definition u has well distributed occurrences (i.e. it has the WELLDOC property), if for any $m \in \mathbb{N}$ and any factor w of u,

$$\{(|Pref_{i_j}u|_0, \ldots, |Pref_{i_j}u|_{d-1}) \bmod m | j \in \mathbb{N}\} = \mathbb{Z}_m^d;$$

that is, the Parikh vectors of $Pref_{i_j}u$ for $j \in \mathbb{N}$, when reduced by modulo m, give the complete set $\mathbb{Z}_m^d$. In order to give the reader a better chance of understanding the WELLDOC property, two examples will be provided.

Example 6. Suppose the Thue-Morse word

$$t = 0110100110010110\ldots$$

is given over binary alphabet, where

$$t_0 = 0,$$

$$t_{2n} = t_n$$

$$t_{2n+1} = 1 - t_n$$

Thue-Morse word does not satisfy the WELLDOC property, if we choose $m = 2$ and $w = 00$. As w occurs only in odd positions i_j, it is easy to see that

$$(|Pref_{i_j}u|_0 + |Pref_{i_j}u|_1) = i_j$$

is odd. It means that

$$(|Pref_{i_j}u|_0, |Pref_{i_j}u|_1) \bmod 2 \neq (0,0),$$

which gives us the necessary

$$\{(|Pref_{i_j}u|_0, |Pref_{i_j}u|_1) \bmod m | j \in \mathbb{N}\} \neq \mathbb{Z}_2^2;$$

Balkova et al. states useful definition and proposition.

Definition 6. We say that an infinite word, u, over an alphabet A, where $|A| = d$, is $universal$ if it contains all finite words over A as its factors.

Proposition 20. *Any word, which is universal, satisfies the WELLDOC property.*

Proof. For any word $w \in A^*$ and any m there exists a finite word v therefore, if $i_0, i_1, \ldots, i_k$ denote the occurrences of w in v, then

$$\{(|Pref_{i_j}u|_0, \ldots, |Pref_{i_j}u|_{d-1}) \bmod m | j \in \{0, 1.\ldots, k\}\} = \mathbb{Z}_m^d.$$

Since u is universal, v is a factor of u. By denoting an occurrence of v in u with i, we get that the positions $i + i_j$ are occurrences of w in u. That gives us

$$\{(|Pref_{i+i_j}u|_0, \ldots, |Pref_{i+i_j}u|_{d-1}) \bmod m | j \in \{0, 1. \ldots, k\}\} =$$
$$= \{(|Pref_i u|_0, \ldots, |Pref_i u|_{d-1})+$$
$$+ \{(|Pref_{i_j}v|_0, \ldots, |Pref_{i_j}v|_{d-1}) \bmod m | j \in \{0, 1. \ldots, k\}\} = \mathbb{Z}_m^d.$$

Thus u satisfies the WELLDOC property. □

Example 7. It is easy to construct a bi-ideal over binary alphabet with WELLDOC property. Suppose we take

$$u_0 = 0, \quad u_1 = 1, \quad u_2 = 00,$$
$$u_3 = 01, \quad u_4 = 10, \quad u_5 = 11,$$
$$u_6 = 000, \quad u_7 = 001, \quad \ldots$$

In other words, we take all the words as generating words in lexicographycal order. In this case our bi-ideal obviously contains all the finite words as its factors, which gives us the WELLDOC property.

3.2 Bounded bi-ideals and WELLDOC property

Theorem 21. *There exists a 1-bounded bi-ideal with the WELLDOC property in a binary alphabet.*

Proof. The idea of the proof is that at first we show how the Parikh vector, when reduced modulo m, gives the whole set $\mathbb{Z}_m^d$ for a fixed factor and a fixed m. Let us choose any factor w of a bi-ideal x and take an arbitrary integer m. From the construction of the bi-ideal there $\exists i \in \mathbb{N}$ (minimal) such that $w \setminus v_i$. Let us define a suffix of v_i, which follows the factor w, by w' (see Figure 1).

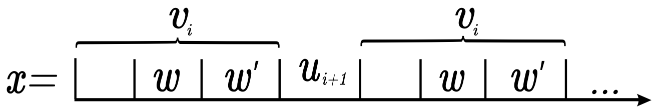

Figure 3.1: Structure of bi-ideal x

Suppose that the first occurrence of w in the bi-ideal x is k. Let us define some integers:

$$p_0 = |Pref_k x|_0 \bmod m, \qquad p_1 = |Pref_k x|_1 \bmod m$$

$$q_0 = |w|_0 \bmod m, \qquad q_1 = |w|_1 \bmod m,$$

$$r_0 = |w'|_0 \bmod m, \qquad r_1 = |w'|_1 \bmod m$$

It is obvious that the Parikh vector for the first occurrence w is (p_0, p_1). Our aim is to show that, at first, we can get the vectors

$$\begin{array}{ccc} (p_0+1, p_1), & (p_0+2, p_1), & \dots, \\ \dots, & (0, p_1), & \dots, \\ \dots, & (p_0-1, p_1), & (p_0, p_1). \end{array}$$

Then we can change the Parikh vector to $(p_0, p_1 + 1)$ and do the same again until we get the whole set $\mathbb{Z}_m^2$.

Further, let us examine v_i which is consecutive $(m+1)$-st and its occurrence which is some integer s_1 (see Figure 2). Let us calculate the Parikh vector for this occurrence w.

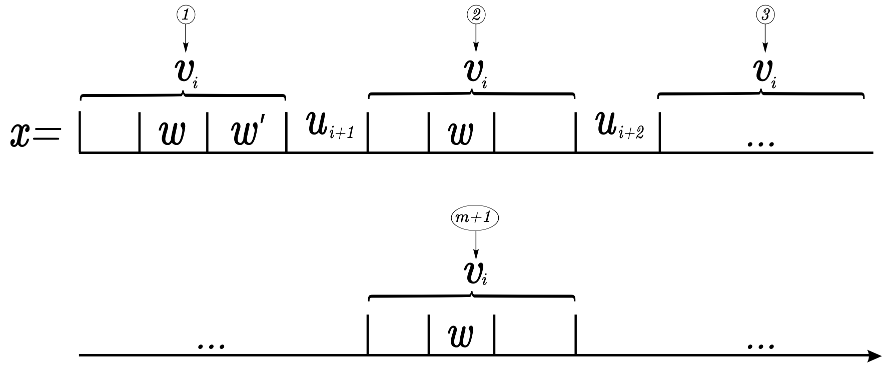

Figure 3.2: Structure of bi-ideal x

$$|Pref_{s_1} v|_0 = p_0 + m(p_0 + q_0 + r_0) + b_0,$$

where b_0 denotes the sum of all the 0's in the base words which comes in the middle of v_i's. Similarly, we get

$$|Pref_{s_1} v|_1 = p_1 + m(p_1 + q_1 + r_1) + b_1,$$

where b_1 denotes the sum of all 1's in the base words which comes in the middle of v_i's. From this we can conclude that

$$(p_0+m(p_0+q_0+r_0)+b_0) \bmod m = (p_0+b_0) \bmod m$$
$$(p_1+m(p_1+q_1+r_1)+b_1) \bmod m = (p_1+b_1) \bmod m.$$

Suppose that the base word with the biggest index in the middle of the issued v_i's is u_{i+j}. Lets choose the base words in such a way:

$$u_{i+l} = \lambda, \forall l \in \{1, 2, \ldots, j-1\},$$

but

$$u_{i+j} = 0.$$

From the construction of the bi-ideal it is known that there is only one base word, u_{i+j}, in the middle. Now we get that

$$(|Pref_{s_1}u|_0, |Pref_{s_1}u|_1) \bmod m = ((p_0+1) \bmod m, p_1).$$

Let us define $k = i+j$ and now observe the element v_k of the bi-ideal sequence and continue to deal with the factor w at occurrence s, which gave us the vector

$$((p_0+1) \bmod m, p_1).$$

Let us again examine that v_k, which is consecutive $(m+1)$-st and its occurrence is some integer s_2 (see figure 3).

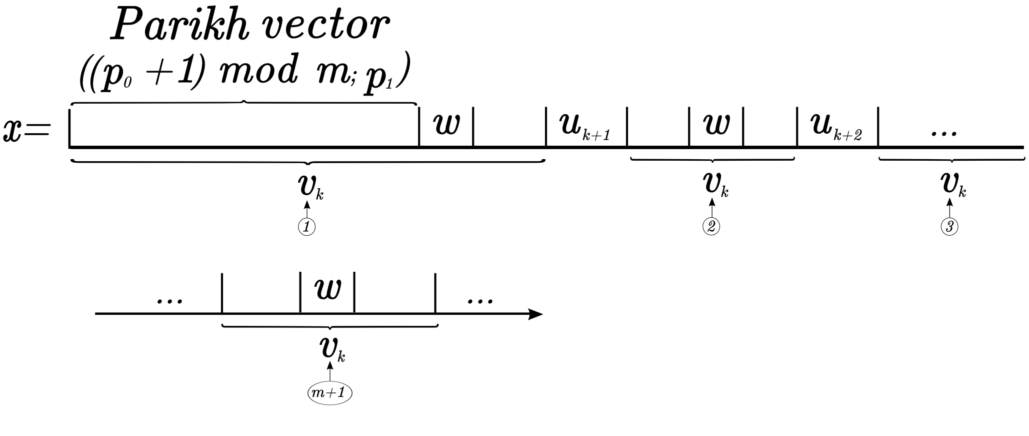

Figure 3.3: Structure of bi-ideal x

We choose the base words again in the following way:

$$u_{k+l} = \lambda, \forall l \in \{1, 2, \ldots, j-1\},$$

but

$$u_{k+j} = 0.$$

Now we can see that

$$(|Pref_{s_2}u|_0, |Pref_{s_2}u|_1) \bmod m = ((p_0 + 2) \bmod m, p_1).$$

Continuing to construct the bi-ideal in such way we get all the vectors

$$(0, p_1), (1, p_1), \dots, (m - 1, p_1).$$

At this point we apply our procedure by taking the base word with the biggest index and defining it by letter 1. This gives us a Parikh vector

$$(p_0, (p_1 + 1) \bmod m).$$

From this point we continue to construct the bi-ideal as before by taking the base word with the biggest index again and defining it by letter 0. Firstly, we get the vector

$$(|Pref_{s'_1}u|_0, |Pref_{s'_1}u|_1) \bmod m = ((p_0 + 1) \bmod m, (p_1 + 1) \bmod m),$$

where s'_1 is the new occurrence of factor w wherewith we are working. By continuing this construction we get the vectors

$$\begin{array}{lll} (0, (p_1 + 1) \bmod m), & (1, (p_1 + 1) \bmod m), & \dots, \\ \dots, & (m - 1, (p_1 + 1) \bmod m). & \end{array}$$

The idea of the further construction is simple. We increase the second coordinate by one (defining the base word with the biggest index by 1) and then forcing in the first coordinate all the possible values reduced modulo m:

$$0,1, \dots, m - 1.$$

After applying this m times the Parikh vectors give the whole set $\mathbb{Z}_m^2$. At this point we have completed the proof for a fixed factor and a fixed integer m.

To make it work for any $m \in \mathbb{N}$ and any factor w of the bi-ideal x, we can make a lexicographic order for all of factors w of x and make an infinite table with all of factors in top row and all the integers (starting from 2, because natural numbers reduced modulo 1 are always 0, so we get vector (0,0) every time and $\mathbb{Z}_1^2 = (0, 0)$) in the left column (see Table 3.2).

To make our bi-ideal with WELLDOC property, we do the following. First, we apply our procedure for the factor w_1 and integer 2 (in the table the cell marked as 1) to get the Parikh

—	w_i	w_1	w_2	w_3	w_4	w_5	w_6
m	—	—	—	—	—	—	—
2	—	1	2	4	7	11	...
3	—	3	5	8	12	...	
4	—	6	9	13	...		
5	—	10	14	...			
6	—	15	...				
7	—	...					

Table 3.1: Sequence of provable cases

vectors to complete the whole set $\mathbb{Z}_1^2$. Then, we take the next marked cell (cell number 2) with a factor w_2 and an integer 2. With the help of the procedure we get the Parikh vectors to complete again the whole set $\mathbb{Z}_1^2$. Then we take the cell number 3 with a factor w_1 and an integer 3, and, again, through the procedure we complete the whole set $\mathbb{Z}_2^2$. By applying our procedure to the all other cells from Table I, we can construct a bi-ideal with the WELLDOC property. Thus, this completes our proof. □

To help the reader understand our construction and procedure, we would like to demonstrate producing a 1-bounded bi-ideal with the WELLDOC property.

Example 8. As we will construct a 1-bounded bi-ideal x, we can arbitrary choose our first base element: $u_0 = 0$. It means that from lexicographical order a factor 0 will be the first one, i.e., $w_1 = 0$. It means that for cell number 1 in Table 1 we have a factor $w_1 = 0$ and an integer $m = 2$. Accordingly, we have to prove that the Parikh vectors for prefixes of factor 0 give the whole set $\mathbb{Z}_2^2$. It results in getting vectors

(0,0), (0,1), (1,0) and (1,1).

We get the vector (0,0) at once because of $Pref_0 x = \lambda$. Thus, the first element of the bi-ideal sequence, v_0, is the one which has $w_1 = 0$ as factor because $v_0 = u_0 = 0$.

As $m = 2$ we have to look at the 3rd of v_0's (see Figure 4).

The base word with the biggest index is u_2. Thus, we define $u_2 = 0$, but $u_1 = \lambda$. From this we have achieved vector (1,0) because

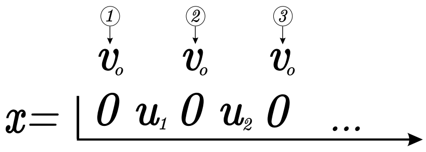

Figure 3.4: Structure of bi-ideal x

$$Pref_3x = 000.$$

Currently we have constructed

$$x = v_2z = 00000z,$$

where z here and further will define a still unknown infinite suffix of the bi-ideal which we are constructing. Further, we have to look at the 3rd of v_2's (see Figure 5).

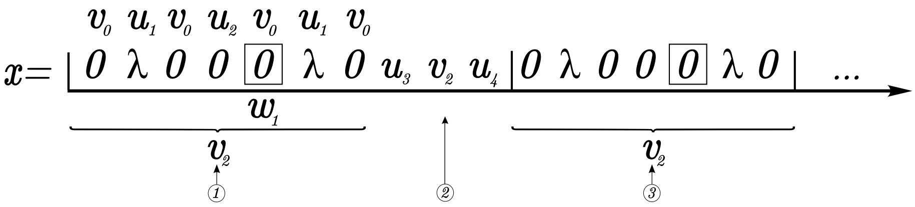

Figure 3.5: Structure of bi-ideal x

The base word with the biggest index is u_4, so we define $u_4 = 0$, but $u_3 = \lambda$. Now we have acquired vector (0,0) back again because

$$Pref_{14}x = 0^{14}.$$

At this point we have constructed

$$x = v_4z = 0^{21}z.$$

Further, we want to increase the second coordinate by one. In order to do that, we have to start including 1's in our bi-ideal. We have to look at the 3rd of v_4's once again (see Figure 6).

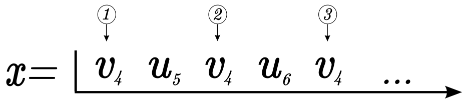

Figure 3.6: Structure of bi-ideal x

The base word with the biggest index is u_6, so we define $u_6 = 1$, but $u_5 = \lambda$. From this we have acquired vector (0,1) back again because

$$Pref_{57}x = 0^{42}10^{14}.$$

For now we have constructed

$$x = v_6z = 0^{42}10^{42}z.$$

To get the whole set $\mathbb{Z}_2^2$ we have to get the Parikh vector (1,1). It means we have to look at the 3rd of v_6's (see Figure 7).

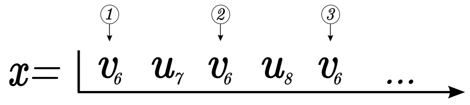

Figure 3.7: Structure of bi-ideal x

The base word with the biggest index is u_8, so we define $u_8 = 0$, but $u_7 = \lambda$. Now we have acquired vector (1,1) back again because

$$Pref_{228}x = 0^{42}10^{42}0^{42}10^{42}00^{42}10^{14}.$$

At this moment we have constructed

$$x = v_8z = v_6u_7v_6u_8v_6u_7v_6 = 0^{42}10^{42}0^{42}10^{42}00^{42}10^{42}0^{42}10^{42}z.$$

At this point we should go to the cell number 2 in Table 1. There we have a factor $w_2 = 1$ and an integer $m = 2$. Further, we have to construct our 1-bounded bi-ideal further so that Parikh vectors for prefixes of factor 1 give the whole set

$$\mathbb{Z}_2^2=\{(0,0),(0,1),(1,0),(1,1)\}.$$

The first occurrence of factor 1 is 42 (known from the element v_8). Thus, we get

$$(|Pref_{42}x|_0, |Pref_{42}x|_1) \text{ mod } 2 = (|0^{42}|_0, |0^{42}|_1) \text{ mod } 2 = (0,0).$$

As $m = 2$ we have to look at the 3rd of v_8's. The base word with the biggest index is u_{10}, so we define $u_{10} = 0$, but $u_9 = \lambda$. In such way we can acquire a vector (1,0). In order to get back vector (0,0) we have to look at the 3rd of v_{10}'s and define $u_{12} = 0$, but $u_{11} = \lambda$. By defining $u_{14} = 1$, but $u_{13} = \lambda$ we get vector (0,1) and by defining $u_{16} = 0$, but $u_{15} = \lambda$ we get vector (1,1). So the Parikh vectors when reduced by modulo 2 give the whole set $\mathbb{Z}_2^2$.

Further we go to the cell number 3 in Table 1 with the factor $w_2 = 0$ and integer $m = 3$. In this case ($m = 3$) the Parikh vectors for prefixes of factor 1 have to give the whole set

$$\mathbb{Z}_3^2 = \{(0,0),(0,1),(0,2),(1,0),(1,1),(1,2),(2,0),(2,1),(2,2)\}$$

The procedure continues to work as before, but at this time it starts with the element v_{16}. Since this time $m = 3$, we have to look at the 4th of v_{16}'s. The base word with the biggest index is u_{18} (in case of $m = 4$ the biggest index increases by 3). Thus, we define $u_{18} = 0$, but $u_{17} = \lambda$, which occours twice this time in the middle.

In general, if we start with element v_i and we have this integer m, we can define

$$b = \lceil log_2(m+1) \rceil.$$

The base word with the biggest index in the middle is u_{i+b} (from the construction of the bi-ideals). Thus, we define $u_{i+b} = 0$ or $u_{i+b} = 1$ (according to the case we want to achieve), but

$$\forall j \in \{0,1,\ldots,b-1\} \colon u_{i+j} = \lambda.$$

Remark 22. *It is easy to see that there are infinitely many 1-bounded bi-ideals with WELLDOC property, because in this construction we can make "the breaks" at some point and include some spontaneous elements to achieve new bi-ideals.*

Remark 23. *We can create a modificated 1-bounded bi-ideal, where some of the first base words are not 1-bounded, but bounded with some $\mu \in \mathbb{N}$.*

For example, we can define first two base words by

$$u_0 = 00110, u_1 = 1101.$$

In this case we can still construct a modified 1-bounded bi-ideal with WELLDOC property by using our construction and procedure.

Proposition 24. *There are infinitely many 1-bounded bi-ideals with WELLDOC property in every finite alphabet A.*

Proof. The idea of this proof is somewhat similar to the proof of the theorem 21. The Parikh vectors have more dimensions (in general case, $|A|$=d). It implies that there will be some more steps in the procedure for the whole set $\mathbb{Z}_m^d$ to be completed. However, it is finite, and the procedure will stop at some point. Although the number of possible factors w_i grows, they are countable and we can arrange them in a row in Table I. □

4 Bounded bi-ideals and linear recurrence

4.1 Preliminaries and background

In this chapter we also give a characterization of linearly recurrent bounded bi-ideals. We introduce the notion of completely bounded bi-ideals and prove that completely bounded bi-ideals are exactly linearly recurrent bounded bi-ideals. This class is very large, namely, its cardinality is continuum.

Theorem 25. *Let w be a word having periods p and q and let $gcd(p,q)$ be the greatest common divisor of p and q. If $|w| \geq p + q - gcd(p,q)$, then w has also the period $gcd(p,q)$.*

Let u be a non-empty factor of $x \in A^\omega$. A word $w \in A^+$ is called *a return word to u of x* if $wu \smallsetminus x$, u is a prefix of wu, and $|wu|_u = 2$. The set of all return words to u of x we denote by $\mathcal{R}_{x,u}$.

If $x \in A^\omega$ is uniformly recurrent, then the difference between two consecutive occurrences of u in x is bounded, therefore $\mathcal{R}_{x,u}$ is finite. As finitely generated bi-ideals and bounded bi-ideals are uniformly recurrent, then $\mathcal{R}_{x,u}$ is finite for each finitely generated (or bounded) bi-ideal x and for each its factor u.

An infinite word $x \in A^\omega$ is called *linearly recurrent* if it is uniformly recurrent and there exists a constant $K \in \mathbb{N}$ such that for all $u \smallsetminus x$ and all $w \in \mathcal{R}_{x,u}$ we have $|w| \leq K \cdot |u|$.

Theorem 26. *Let x be an aperiodic linearly recurrent word with constant K. Then:*

1. *For all $n \in \mathbb{N}$ each subword of length n appears in each factor of length $(K+1)n$ in x.*

2. *The number of distinct factors of length n in x is less than or equal to Kn.*

3. *For all $u \in \mathbf{F}(x)$ and for all $w \in \mathcal{R}_{x,u}$ we have $(1/K)|u| < |w|$.*

4. *For all $u \in \mathbf{F}(x)$, Card$(\mathcal{R}_{x,u}) \leq K(K+1)^2$.*

In (Durand, 2000) and (Durand, 2003) Durand gave a S-adic characterization of linearly recurrent sequences. If S is a set of morphisms (possibly infinite), an *S-adic representation* of x is given by a sequence $(\sigma_n : A_{n+1}^* \to A_n^*)_{n\in\mathbb{N}}$ of morphisms in S and a sequence $(a_i)_{i\in\mathbb{N}}$ of letters ($a_i \in A_i$ for all $i \in \mathbb{N}$) such that $A = A_0$, $x = \lim_{n\to+\infty} \sigma_0\sigma_1 \ldots \sigma_n(a_{n+1}^\omega)$ and $\lim_{n\to+\infty} |\sigma_0\sigma_1 \ldots \sigma_n(a_{n+1})| = +\infty$. An infinite word x over alphabet A is called S-adic if there exists a set S of morphisms such that x admits S-adic representation. If there exists $s \in \mathbb{N}$ such that for all $r \in \mathbb{N}$, for all $b \in A_r$, and $c \in A_{r+s+1}$ the letter b occurs in $\sigma_{r+1}\sigma_{r+2} \ldots \sigma_{r+s}(c)$, then x is called a *primitive* S-adic sequence (with constant s). A morphism $\sigma : A \to B^*$ is called *proper* if there exist two letters $r, l \in B$ such that for all $a \in A$ the first letter of $\sigma(a)$ is l, and the last letter of $\sigma(a)$ is r. We say that $x \in A^\omega$ is *proper S-adic* if it is S-adic, and all morphisms $\sigma \in S$ are proper.

Proposition 27. *A sequence is linearly recurrent if and only if it is a primitive and proper S-adic sequence.*

According to Proposition 27 in order to check whether a bounded bi-ideal x is linearly recurrent or is not linearly recurrent we should be able to choose a finite set S of proper morphisms and show that x is primitive and proper S-adic, or to prove that there does not exist a finite set of proper morphisms S such that x is primitive and proper S-adic. For our purposes it is more convenient to make restrictions on the basis sequence (u_i) before the bi-ideal is generated.

4.2 Bounded Bi-ideals and Linear Recurrence

In this section we state and prove the main result of this chapter, that is, completely bounded bi-ideals are exactly linearly recurrent bounded bi-ideals.

4.2.1 Completely Bounded Bi-ideals

Let

$$u_0, u_1, \ldots, u_n, \ldots \tag{4.21}$$

be a sequence of finite words over alphabet A.

A subsequence of sequence (4.21)

$$u_{i_0}, u_{i_1}, \ldots, u_{i_k}, \ldots \tag{4.22}$$

is called *constant* if the following conditions hold:

(i) $\forall j, j' \in \mathbb{N} \left(u_{i_j} = u_{i_{j'}} \right)$;

(ii) $u_k = u_{i_0} \implies \exists n(k = i_n)$.

A constant subsequence (4.22) is called *bounded* if it is finite or there exists a positive integer l such that

$$i_n - i_{n-1} \leq l$$

for all $n \in \mathbb{N}$. A constant subsequence (4.22) is called *boundless* if it is not bounded.

Definition 7. A bounded bi-ideal x is called *completely bounded* if there exists a basis sequence (4.21) of x which contains only bounded constant subsequences.

Since each bi-ideal has infinitely many basis sequences, then the "existence" condition in Definition 7 is crucial.

Example 9. Let x be a periodic word $(01)^\omega$. It is a completely bounded bi-ideal since its basis sequence

$$01, 01, 01, \ldots, 01, \ldots$$

has only one constant subsequence, and it is bounded. Nevertheless, the sequence $(u_i)_{i \geq 0}$ which is defined by

$$u_i = \begin{cases} 01 & \text{if } i = k^2 \text{ for a } k \in \mathbb{N}, \\ 0101 & \text{otherwise} \end{cases}$$

also is a basis sequence of x. Clearly, $(u_i)_{i \geq 0}$ contains a boundless constant subsequence.

Let x be a bounded bi-ideal generated by a sequence (4.21). As the length of each basis word of x is bounded by some $s \in \mathbb{N}$, the set

$$\mathcal{U} = \{u \big| \, \exists k(u = u_k)\}$$

is finite. From here there exists a non-negative integer m such that

$$\mathcal{U} = \{u \big| \, \exists k \in \overline{0, m}(u = u_k)\}. \tag{4.23}$$

We denote $\mu_0 = \min\limits_{u \in \mathcal{U}} |u|$, and $\mu_1 = \max\limits_{u \in \mathcal{U}} |u|$.

Let (u_n) be a basis sequence of a completely bounded bi-ideal x such that all constant subsequences (4.22) of (u_n) are bounded. Then there exists a positive integer l such that for all constant subsequences (4.22) of the basis sequence (u_n) we have

$$i_n - i_{n-1} \leq l \tag{4.24}$$

for all $n \in \mathbb{N}$.

In the sequel we use denotation m (l, respectively) for the smallest integer that satisfies (4.23) ((4.24), respectively) for a completely bounded bi-ideal with a given basis sequence (u_n).

We recall that by v_n we denote the n-th element of the bi-ideal sequence that is generated by a basis sequence (u_n), i.e., $v_0 = u_0$ and $v_{n+1} = v_n u_{n+1} v_n$ for all $n \in \mathbb{N}$.

Lemma 28. *Let x be a bounded bi-ideal with a basis sequence (u_n). Let (v_n) be a bi-ideal sequence generated by (u_n). Then $(2^{k+1}-1)\mu_0 \leq |v_k| \leq (2^{k+1}-1)\mu_1$ for all $k \in \mathbb{N}$.*

Proof. If $k = 0$, then $v_0 = u_0$, therefore

$$\mu_0 \leq |u_0| \leq \mu_1.$$

Now we assume that

$$(2^{k+1}-1)\mu_0 \leq |v_k| \leq (2^{k+1}-1)\mu_1$$

and consider the length of v_{k+1}. The equality

$$|v_{k+1}| = 2 \cdot |v_k| + |u_{k+1}|$$

implies

$$|v_{k+1}| \geq 2(2^{k+1}-1)\mu_0 + \mu_0 = (2^{k+2}-1)\mu_0$$

and

$$|v_{k+1}| \leq 2(2^{k+1}-1)\mu_1 + \mu_1 = (2^{k+2}-1)\mu_1.$$

□

Lemma 29. *Let x be a bounded bi-ideal with a basis sequence (u_n). Let (v_n) be a bi-ideal sequence generated by (u_n). Then $|v_{k+n}| \leq 2^n|v_k| + (2^n-1)\mu_1$ for all $k, n \in \mathbb{N}$.*

Proof. The proof is by induction. If $n = 1$, then

$$|v_{k+1}| = 2 \cdot |v_k| + |u_{k+1}| \leq 2 \cdot |v_k| + (2-1)\mu_1.$$

We assume that condition holds for $|v_{k+n}|$ and consider the length of v_{k+n+1}:

$$\begin{aligned}|v_{k+n+1}| &= 2 \cdot |v_{k+n}| + |u_{k+n+1}| \leq 2 \cdot (2^n|v_k| + (2^n-1)\mu_1) + \mu_1 \\ &= 2^{n+1}|v_k| + (2^{n+1}-1)\mu_1.\end{aligned}$$

□

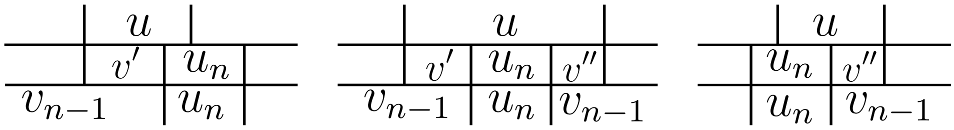

Figure 4.1: Three possibilities where the word u can occur in $v'u_nv''$.

From now we will consider only completely bounded bi-ideals.

Lemma 30. *Let x be a completely bounded bi-ideal. Let (u_n) be a basis sequence of x that contains only bounded constant subsequences. Let (v_n) be a bi-ideal sequence generated by the sequence (u_n). If $u \in \mathbf{F}(v_n)$, $u \notin \mathbf{F}(v_{n-1})$, and $n \geq m+1$, then $|u| > |v_{n-1-l}|$.*

Proof. Firstly, we observe that $v_n = v_{n-1}u_nv_{n-1}$ and $u \notin \mathbf{F}(v_{n-1})$ imply $u \searrow v'u_nv''$, where $v' \in \text{Pref}(u) \cap \text{Suff}(v_{n-1})$ and $v'' \in \text{Suff}(u) \cap \text{Pref}(v_{n-1})$. We can represent this condition with three alternative schemes (see Figure 4.1).

As one can see, it is possible to have $v' = \lambda$ or $v'' = \lambda$.

Definition of a completely bounded bi-ideal implies existence of i, $1 \leq i \leq l$, such that $u_n = u_{n-i}$, but from bi-ideal construction we have $v_{n-1-i} \in \text{Pref}(v_{n-1})$ and $v_{n-1-i} \in \text{Suff}(v_{n-1})$. From here we obtain

$$v_n = v_{n-1}u_nv_{n-1} = w_1v_{n-i}w_2,$$

with $|w_1| = |w_2|$. For a schematical representation see Figure 4.2.

Now, one can see that the inequality $|u| \leq |v_{n-1-i}|$ implies $|v'| \leq |v_{n-1-i}|$ and $|v''| \leq |v_{n-1-i}|$.

Thus $v'u_nv'' \searrow v_{n-i} \in \text{Pref}(v_{n-1})$. Contradiction, since $u \notin \mathbf{F}(v_{n-1})$. □

Corollary 31. *Let x be a completely bounded bi-ideal. Let (u_n) be a basis sequence of x that contains only bounded constant subsequences. Let (v_n) be a bi-ideal sequence generated by (u_n). If $u \searrow v_n$, but u does not appear in v_{n-1}, then $|v_n| < 2^{\varkappa}(|u| + \mu_1)$, where $\varkappa = \max\{m+1, l+1\}$.*

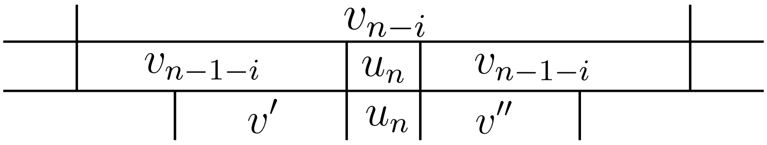

Figure 4.2: The structure of v_{n-i}.

Proof. (i) Let $n \geq m+1$. Then from Lemma 29 and Lemma 30 we obtain

$$\begin{aligned}|v_n| &\leq 2^{l+1}|v_{n-1-l}| + (2^{l+1}-1)\mu_1 \\ &< 2^{l+1}|v_{n-1-l}| + 2^{l+1}\mu_1 \\ &< 2^{l+1}|u| + 2^{l+1}\mu_1 \\ &= 2^{l+1}\left(|u| + \mu_1\right) \\ &\leq 2^{\varkappa}\left(|u| + \mu_1\right).\end{aligned}$$

(ii) If $n \leq m$, then by Lemma 29 it follows that

$$\begin{aligned}|v_n| &\leq 2^n|v_0| + (2^n-1)\mu_1 < 2^n|u_0| + 2^n\mu_1 \\ &\leq 2^n\mu_1 + 2^n\mu_1 \leq 2^{m+1}\mu_1 \leq 2^{m+1}\left(|u| + \mu_1\right) \\ &\leq 2^{\varkappa}\left(|u| + \mu_1\right).\end{aligned}$$

□

Lemma 32. *If $a \geq 1$ and $b \geq 1$, then $a+b \leq ab+1$.*

Proof. From inequalities $a \geq 1$ and $b \geq 1$ we easily obtain

$$\begin{aligned}a-1 &\leq (a-1)b, \\ a-1 &\leq ab-b, \\ a+b &\leq ab+1.\end{aligned}$$

□

Corollary 33. *If $a \geq 1$ and $b \geq 1$, then $a+b \leq 2ab$.*

Proof. Lemma 32 implies

$$a+b \leq ab+1 \leq ab+ab = 2ab.$$

□

Corollary 34. *Let x be a completely bounded bi-ideal. Let (u_n) be a basis sequence of x that contains only bounded constant subsequences. Let (v_n) be a bi-ideal sequence generated by (u_n). If $u \searrow v_n$ and $u \notin \mathbf{F}(v_{n-1})$, then*

$$|v_n| < 2^{\varkappa+1}\mu_1|u|,$$

where $\varkappa = \max\{m+1, l+1\}$.

Proof. Corollary 31 and Corollary 33. □

4.2.2 The Main Result

Theorem 35. *A bounded bi-ideal x is linearly recurrent if and only if it is completely bounded.*

Proof. $\Leftarrow$: At first we prove that a completely bounded bi-ideal x is linearly recurrent. Let x be a completely bounded bi-ideal, and $(u_i)_{i\geq 0}$ be its basis sequence that contains only bounded constant subsequences. Let $u \searrow x$. Then there exists an element v_n of the bi-ideal sequence such that $u \searrow v_n$. By construction of a completely bounded bi-ideal, x can be written as a factorization of v_n and basis words, i.e.,

$$x = v_n u'_1 v_n u'_2 \dots v_n u'_k \dots,$$

where $u'_s \in \mathcal{U}$ for all $s \in \mathbb{N}_+$.

Let $u[\mathbf{i}, \mathbf{j})$ be an occurrence of u in x such that $u = x[i, j)$. Then there is $k \in \mathbb{N}$ and occurrence of $v_n u'_k v_n$ in x

$$v_n[\mathbf{i_1}, \mathbf{i_2}) u'_k[\mathbf{i_2}, \mathbf{i_3}) v_n[\mathbf{i_3}, \mathbf{i_4}) = x[i_1, i_4)$$

such that $i_1 \leq i < i_3$. Otherwise, for $i \in [i_3, i_4)$ we would consider the occurrence of $v_n u'_{k+1} v_n$ in x instead of $v_n u'_k v_n$. So, $i_1 \leq i < i_3$ and we will find the next occurrence of u in x, e.g., $u[\mathbf{i'}, \mathbf{j'}) = x[i', j')$. As u has an occurrence in v_n, then $i_3 \leq i' < i_4$. Clearly, $u[\mathbf{i}, \mathbf{j})$ and $u[\mathbf{i'}, \mathbf{j'})$ are two distinct occurrences of u in x and we can estimate the length of w, e.g., the length of the corresponding return word to u is

$$|w| \leq i' - i \leq |v_n u'_k| \leq |v_n| + \mu_1.$$

(i) If u does not appear in v_{n-1}, then, by Corollary 34, we have

$$|w| \leq |v_n| + \mu_1 < 2^{\varkappa+1}\mu_1|u| + \mu_1 \leq 2^{\varkappa+2}\mu_1|u|.$$

(ii) Observe, if $u \searrow v_{n-1}$, then we need to consider only the case when $u \searrow v_0$. Then

$$|w| \leq |v_0| + \mu_1 \leq 2\mu_1 \leq 2^{\varkappa+2}\mu_1|u|.$$

We conclude the proof by setting $K = 2^{\varkappa+2}\mu_1$. Then for each $u \searrow x$ and each return word $w \in \mathcal{R}_{x,u}$ we have

$$|w| \leq K \cdot |u|.$$

$\Rightarrow$: We assume the contrary that x is a linearly recurrent bounded bi-ideal that is not completely bounded. Then there exists a constant K such that for each factor w we have

$$|w| \leq K \cdot |r_w|,$$

where r_w is arbitrary return word to w in a bi-ideal x.

Let (u_i) be a basis sequence of x. Without loss of generality we can assume that the length μ_0 of the shortest basis word is greater than zero. Indeed, if $\mu_0 = 0$ (i.e., at least one of the basis words is the empty word λ), then we can L-prolong the basis words once and consider the new obtained basis sequence $(u_i^{(1)})$ instead of (u_i). By definition of L-prolongation, $u_0 \neq \lambda$ is a prefix of $u_i^{(1)}$ for all $i \in \mathbb{N}$. Thus none of the basis words in $(u_i^{(1)})$ is empty. Clearly, as x is not a completely bounded bi-ideal, then $(u_i^{(1)})$ contains a boundless constant subsequence.

Let $k_i = |u_i| - |u_{i+1}|$ for all $i \in \mathbb{N}$. Let $\mathcal{K} = \{|k_i| | i \in \mathbb{N}\}$. According to the definition of L-prolongation

$$k_i^{(1)} = |u_i^{(1)}| - |u_{i+1}^{(1)}| = |u_0 u_{i+1}| - |u_0 u_{i+2}| = k_{i+1}.$$

Hence, if $\mathcal{K}' = \{|k_i^{(1)}| | i \in \mathbb{N}\}$, then $\mathcal{K}' \subset \mathcal{K}$. From here, without loss of generality we can assume that the basis sequence (u_i) satisfies such two conditions

$$\mu_1 < 2\mu_0, \tag{4.25}$$

where μ_0 (μ_1, respectively) is the length of the shortest (longest, respectively) basis word, and

$$0 < k_{\text{max}} < 0.1\mu_0, \tag{4.26}$$

where $k_{\text{max}} = \max\{||u_i| - |u_{i+1}||\,;\; i \in \mathbb{N}\}$.

Otherwise we could L-prolong basis words until we obtain a basis sequence that satisfies (4.25) and (4.26). Since x is not a completely bounded bi-ideal, then after any number of L-prolongations we obtain a basis sequence that contains a boundless constant subsequence.

The aperiodicity of x implies the existence of a constant n' such that for all $n'' \geq n'$ the n''-th element of the bi-ideal sequence $v_{n''}$ is not p-periodic for all $p \leq 3\mu_1$. Later this fact will help us to obtain a contradiction.

As the basis sequence of x contains a boundless constant subsequence (u_{i_k}), then there exists $n > \max\{2^m, n'\}$ such that $n = i_k$, $u_{i_k} = u_{i_{k+1}}$, and $i_{k+1} - i_k \geq K + 3$. In order to simplify calculations we assume that equality holds, namely, $i_{k+1} = n + K + 3$. We prove that for a word $w = v_n u_n v_n$ there exists a return word r_w such that $|r_w| > K \cdot |w|$.

Further we divide the proof in two parts. First we prove that v_{n+K+2} does not contain $w = v_n u_n v_n$ as a factor. Then we show that if w does not occur in v_{n+K+2}, then there exists a return word to w such that its length is greater than $K \cdot |w|$.

By the bi-ideal construction, it follows that v_{n+K+2} can be written as a factorization of v_n and basis words. Hence, if we denote each $u_i \in \mathcal{U} \setminus \{u_n\}$ by u_*, then v_{n+K+2} can be written in

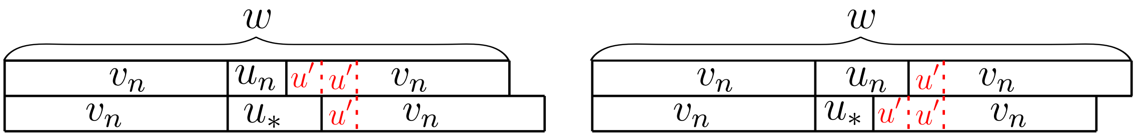

Figure 4.3: The word w is in the prefix of v_{n+K+2}.

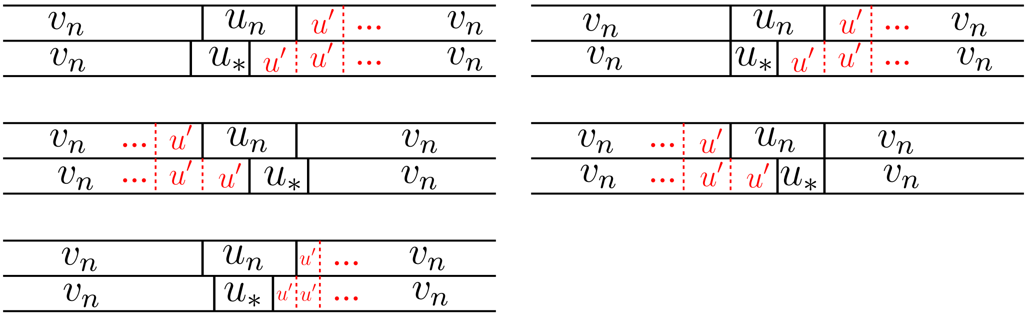

Figure 4.4: The word u_n overlaps with u_*; the case when $|u_n| > |u_*|$.

the form

$$v_{n+K+2} = v_n u_* v_n u_* v_n u_* \dots v_n u_* v_n. \tag{4.27}$$

We use denotation u_* to point out the absence of u_n in factorization (4.27) of v_{n+K+2}, i.e., to point out the inequality $u_* \neq u_n$.

Firstly, we observe that $w \notin \text{Pref}(v_{n+K+2})$. Assume the contrary that $w \in \text{Pref}(v_{n+K+2})$. Then $|u_n| \neq |u_*|$ (otherwise the equality $u_* = u_n$ leads to contradiction). Hence we obtain a shift of v_n to the right of u_n and u_* (see Figure 4.3); therefore v_n is periodic with the length of the shift $|u'| < \mu_1 < 3\mu_1$. Contradiction. Analogously, we can prove that $w \notin \text{Suff}(v_{n+K+2})$.

Next, we consider the case when $u_n = w[|v_n|, |v_n u_n|)$ occurs in $v_n u_* v_n u_* v_n$ so that u_n overlaps with one of the basis words u_*. As $u_n \neq u_*$, then these two conditions cannot hold at the same time:

a) $|u_*| = |u_n|$;

b) occurrences of words u_* and u_n in v_{n+K+2} start at the same position.

From here we obtain the shift of v_n to the right or to the left of u_n and u_* (see Figure 4.4), therefore v_n is periodic with the length of the shift $|u'| < \mu_1 < 3\mu_1$. Contradiction.

It remains to consider the case when $u_n = w[|v_n|, |v_n u_n|)$ occurs in

$$v_n = (v_n u_* v_n u_* v_n)[|v_n u_*|, |v_n u_* v_n|).$$

Here we recall that $v_n = v_{n-1} u_n v_{n-1}$ and consider three subcases:

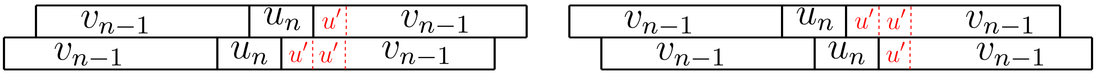

Figure 4.5: There is a shift, when u_n overlaps with u_n in v_n.

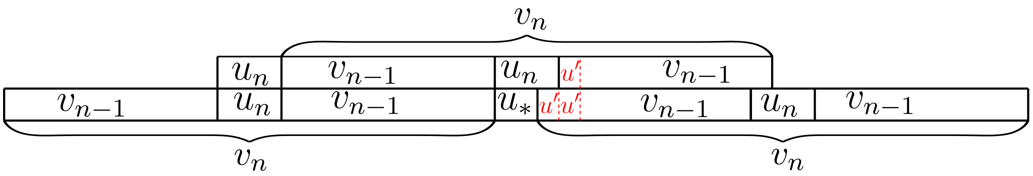

Figure 4.6: The case when $|u_n| > |u_*|$.

Case I

The word $u_n = w[|v_n|, |v_n u_n|)$ overlaps with $u_n = v_n[|v_{n-1}|, |v_{n-1}u_n|)$ so that

a) there is a shift of v_{n-1} to the left or to the right of u_n (see Figure 4.5). Then v_{n-1} is periodic with the length of the shift $|u'| < \mu_1 < 3\mu_1$. Contradiction.

b) their occurrences coincide, namely,

$$w = v_n u_n v_n = v_{n-1}u_n v_{n-1}u_n v_{n-1}u_n v_{n-1} =$$

$$= (v_n u_* v_{n-1} u_n v_{n-1} u_* v_n)[|v_n u_*| - |v_{n-1}u_n|, |v_n u_* v_n| + |u_n v_{n-1}|].$$

From here the equality of lengths of u_n and u_* would imply the equality of words themselves (which would lead to contradiction). Thus $|u_n| \neq |u_*|$ and we obtain shift of v_{n-1} to the right of u_* and u_n (see Figure 4.6).

From here v_{n-1} is periodic with the length of the shift $|u'| < k_{\text{max}} < \mu_1 < 3\mu_1$. Contradiction.

Case II

Let $v_{n-1} = v'v''v'''$ with $|v'| = |v'''| = 2\mu_1$. Let $u_n = w[|v_n|, |v_n u_n|)$ occur in $v' = v_{n-1}[0, 2\mu_1)$ or $v''' = v_{n-1}[|v_{n-1}| - 2\mu_1, |v_{n-1}|)$. Here we are not interested if u_n occurs in $v_{n-1} \in \text{Pref}(v_n)$ or in $v_{n-1} \in \text{Suff}(v_n)$. If u_n occurs in v' starting at position α, then, as

$$w[|v_n|, |v_n u_n v_{n-1}|) = u_n v_{n-1},$$

we obtain a shift of v_{n-1} of length $\alpha + |u_n|$ to the right of u_n (see Figure 4.7). Hence v_{n-1} is periodic with the length of the shift

$$\alpha + |u_n| \leq |v'| = 2\mu_1.$$

Contradiction.

Similarly, if u_n occurs in v''' ending in position $|v_{n-1}| - \beta$, then, since

$$w[|v_{n-1}u_n|, |v_n u_n|) = v_{n-1}u_n,$$

we have shift of v_{n-1} of length $\beta + |u_n|$ to the left of u_n (see Figure 4.7), therefore v_{n-1} is periodic with a period

$$\beta + |u_n| \leq |v'''| = 2\mu_1.$$

Contradiction.

Case III

Let $v_{n-1} = v'v''v'''$, where $|v'| = |v'''| > \mu_1$. Let $u_n = w[|v_n|, |v_n u_n|)$ occur in $v'' = v_{n-1}[|v'|, |v'v''|)$. First we consider the case, when u_n occurs in v'' which is a factor of $v_{n-1} \in$ Pref(v_n). If $u_n = w[|v_n|, |v_n u_n|)$ occurs in v_{n-1} starting in position α (see Figure 4.8), then v_n is both $\alpha + |u_*|$ and $\alpha + |u_n|$ periodic.

a) If $|u_*| \neq |u_n|$, then according to Theorem 25: if

$$\begin{aligned}&\alpha + |u_*| + \alpha + |u_n| - gcd(\alpha + |u_*|, \alpha + |u_n|)\\ =&2\alpha + |u_n| + |u_*| - gcd(\alpha + |u_*|, \alpha + |u_n|) < |v_n|,\end{aligned}$$

then v_n is also $gcd(\alpha + |u_*|, \alpha + |u_n|)$ periodic. Indeed, we have

$$2\alpha + |u_n| + |u_*| - gcd(\alpha + |u_*|, \alpha + |u_n|) < 2|v_{n-1}| - 2\mu_1 + 2\mu_1 - 1 < |v_n|,$$

therefore v_n is

$$\begin{aligned}&gcd(\alpha + |u_*|, \alpha + |u_n|)\\ &= gcd(\min(\alpha + |u_*|, \alpha + |u_n|), ||u_*| - |u_n||) \leq ||u_*| - |u_n|| < \mu_1\end{aligned}$$

periodic. Contradiction.

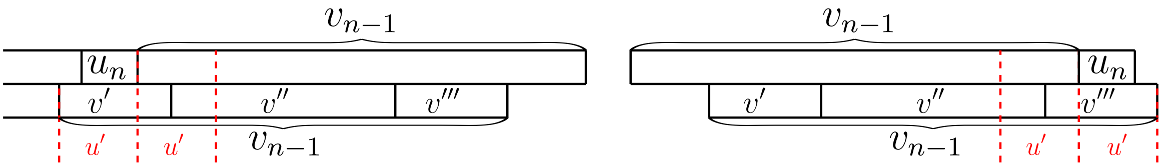

Figure 4.7: The word u_n occurs in v' or v'''.

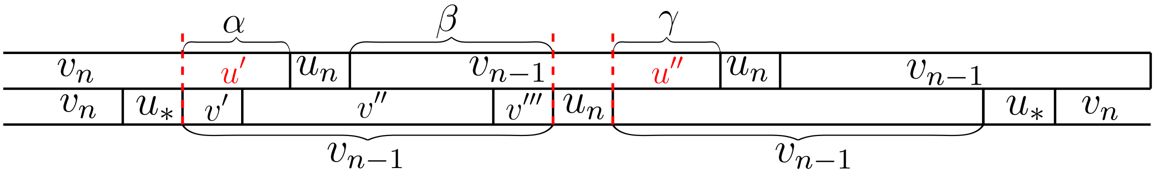

Figure 4.8: The word u_n occurs in v''.

b) If $|u_*| = |u_n|$, then we consider u' and u'' (see Figure 4.8), two suffixes of v_{n-1}, and compare their lengths α and γ:

$$\alpha = |u'| = |v_{n-1}| - \beta - |u_n| = \gamma.$$

Hence u_*u' and u_nu'' are two suffixes of v_{n-1} of the same length. From here and the equality of lengths of u_n and u_* we obtain that

$$u_n = u_*$$

as prefixes of the same length of equal words. Contradiction.

If $u_n = w[|v_n|, |v_nu_n|)$ occurs in $v_{n-1} \in \mathrm{Suff}(v_n)$ ending in position $|v_{n-1}| - \alpha$ (see Figure 4.9), then v_n is both $\alpha + |u_*|$ and $\alpha + |u_n|$ periodic. Analogously as before we obtain:

a) If $|u_*| \neq |u_n|$, then v_n is also

$$gcd(\alpha + |u_*|, \alpha + |u_n|) < \mu_1$$

periodic, which leads to contradiction.

b) If $|u_*| = |u_n|$, then $|u''| = |u'|$ (see Figure 4.9), hence $u'' = u'$ as prefixes of v_{n-1} of equal length. From here $u_* = u_n$ as suffixes of equal length of the same word $u''u_n = u'u_* \in \mathrm{Pref}(v_{n-1})$.

We have proved that $w = v_nu_nv_n$ does not occur in v_{n+K+2}. It remains to show that this implies existence of a return word to w such that its length exceeds $K \cdot |w|$. As v_{n+K+2} does

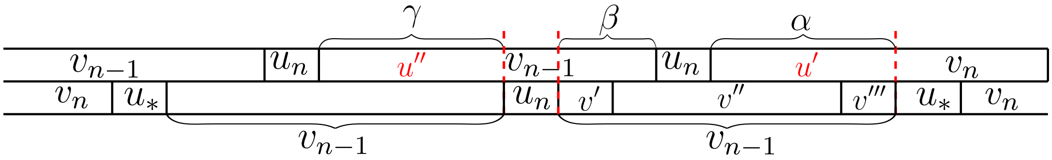

Figure 4.9: The word u_n occurs in v''.

not contain w as a factor, there exists a return word r_w to w such that $r_w w$ contains v_{n+K+2} as a factor. We need to prolong v_{n+K+2} for at least two symbols to obtain $r_w w$ (if we add one symbol to the right or to the left, we can obtain maximum one occurrence of w). From here

$$|r_w| = |r_w w| - |w| \geq |v_{n+K+2}| + 2 - |w|. \tag{4.28}$$

Now we estimate the length of w. By Lemma 28:

$$|w| = |v_n u_n v_n| \leq 2|v_n| + \mu_1 \leq 2(2^{n+1} - 1)\mu_1 + \mu_1 < 2^{n+2}\mu_1. \tag{4.29}$$

Next, from (4.28), (4.29), and Lemma 28 we have

$$\begin{aligned}
|r_w| &\geq |v_{n+K+2}| + 2 - |w| \\
&> (2^{n+K+3} - 1)\mu_0 + 2 - 2^{n+2}\mu_1 \\
&= 2^{n+K+3}\mu_0 - \mu_0 + 2 - 2^{n+2}\mu_1 \\
&> 2^{n+K+2}\mu_1 - 2^{n+1}\mu_1 - 2^{n+2}\mu_1 \\
&= \left(2^K - \frac{3}{2}\right) \cdot 2^{n+2}\mu_1 \\
&> \left(2^K - \frac{3}{2}\right) \cdot |w|.
\end{aligned}$$

Finally, we conclude the proof by observing that for each integer $K > 1$ we have $2^K - \frac{3}{2} > K$. Hence

$$|r_w| > K \cdot |w|,$$

and x is not LR. Contradiction. □

We have given a characterization of linearly recurrent bounded bi-ideals. Moreover, the famous Thue-Morse word is linearly recurrent (as uniformly recurrent morphic infinite word), but it is not a bounded bi-ideal (Buls and Lorencs, 2006). Hence we conclude that class of bounded bi-ideals intersects with the class of linearly recurrent words but neither of these classes is a proper subclass of another one (see Figure 4.10).

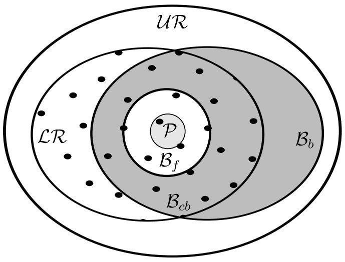

Figure 4.10: Hierarchy of uniformly recurrent words ($\mathcal{UR}$): the class of linearly recurrent words – $\mathcal{LR}$, the class of bounded bi-ideals – $\mathcal{B}_b$, the class of completely bounded bi-ideals – $\mathcal{B}_{cb}$, the class of finitely generated bi-ideals – $\mathcal{B}_f$, the class of periodic words – $\mathcal{P}$.

5 Partial finitely generated bi-ideals

5.1 Preliminaries and background

Blanchet-Sadri et al. in (Blanchet-Sadri and Hegstrom, 2002) accented that partial words appear naturally in several fields such as DNA computing, data communication, molecular biology etc. This was the inspiration of this chapter – to aggregate partial words with the class of infinite words that we are interested in, i.e., bi-ideals. Nowadays the importance of information is so expansive that it is not possible to overvalue it, as there are times and reasons for knowing only partial information about something, for example, DNA structure. As DNA has some certain structure (with possible missing information) and bi-ideals (in this case, finitely generated bi-ideals) have a structure, in this chapter we are trying to solve the problem of filling the holes (missing information) in finitely generated bi-ideals. In general case of finite amount of holes in finitely generated bi-ideals it is always possible to get the whole information back. Unfortunately, in general case of infinite amount of holes it is not possible.

A *finite partial word* of length n over A is a map $w : \{0, \dots, n-1\} \to A \cup \{\diamond\}$, where $\diamond \notin A$. The symbol $\diamond$ is viewed as a "do not know" symbol. The union set $A \cup \{\diamond\}$ is denoted by $A_\diamond$. A *right infinite partial word* or *infinite partial word* over A is a map $w : \mathbb{N} \to A_\diamond$. In both the finite and infinite cases, the symbol at position i in w is denoted by w_i. If $w_i \in A$, then i is defined in w, and if $w_i = \diamond$, then i is a hole in w.

5.2 Finding a basis for a bi-ideal

We start this section with the theorem what gives an orientation in this chapter.

Theorem 36. *Suppose that $\langle u_0, u_1, ..., u_{m-1}\rangle$ is a basis that generates the bi-ideal x. If $u'_0 \in Pref(x)$ and $u'_0 \neq u_0$, then there exists a basis $\langle u'_0, u'_1, ..., u'_{m-1}, u'_m, u'_1, u'_2, ..., u'_m, u'_1, ...\rangle$ for almost finitely generated bi-ideal, which generates the same bi-ideal x.*

Proof. From given it is known that $u_0' \in Pref(x)$. It follows that either $u_0' \in Pref(v_0)$ or either $\exists n > 0 : u_0' \in Pref(v_n)$ and $u_0' \notin Pref(v_{n-1})$. It means that $\exists w$:

$$v_n = u_0'w \tag{5.21}$$

From the bi-ideal construction and (5.21), we have

$$v_{n+1} = v_n u_{n+1} v_n = u_0' w u_{n+1} u_0' w.$$

Let's define $u_1' := wu_{n+1}$ which gives us

$$v_{n+1} = u_0'u_1'u_0'w = v_0'u_1'v_0'w = v_1'w.$$

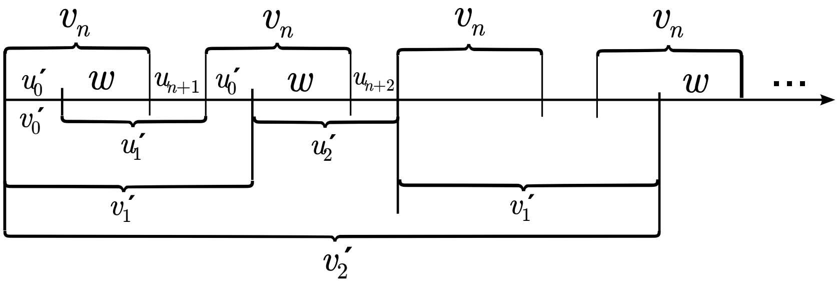

Figure 5.1: Scheme of change of basis

Let's examine the next element of the bi-ideal sequence:

$$v_{n+2} = v_{n+1}u_{n+2}v_{n+1} = v_1'wu_{n+2}v_1'w.$$

Let's define again similarly $u_2' := wu_{n+2}$ and it gives us

$$v_{n+2} = v_1'u_2'v_1'w = v_2'w.$$

We will use mathematical induction to prove this theorem. Suppose that $v_{n+i} = v_i'w$, where $i < m$. It is easy to see that

$$v_{n+i+1} = v_{n+i}u_{n+i+1}v_{n+i} = v_i'wu_{n+i+1}v_i'w.$$

If we define $u_{i+1}' := wu_{n+i+1}$, we get

$$v_{n+i+1} = v_{i+1}'w.$$

Now we have to make an inductive step, i.e., we have to get $v_{n+m+1} = v_{m+1}'w$. From the bi-ideal construction and from the fact that the initial basis of finitely generated bi-ideal is a basis with m elements we get that

$$v_{n+m+1} = v_{n+m}u_{n+m+1}v_{n+m} = v_{n+m}u_{n+1}v_{n+m} = v_m'wu_{n+1}v_m'w.$$

Bus as we know from the definition that $u'_1 := wu_{n+1}$, then

$$v_{n+m+1} = v'_m u'_1 v'_m w = v'_{m+1} w.$$

□

Example 10. Let's observe a basis $\langle u_0, u_1 \rangle$, where $u_0 = 0$ and $u_1 = 1$. This basis generates a bi-ideal $x = 0100010101000100...$ Suppose that we choose $u'_0 = 01000$ and try to get a basis for an almost finitely generated bi-ideal, which generates the same bi-ideal x. As $u'_0 = 01000$, then it is known that $u'_0 \in Pref(v_2)$. It means that there exists a word $w = 10$ such that:

$$v_2 = u'_0 w$$

Inductive hypothesis: $\forall i \leq k$: $v_{i+2} = v'_i w$. Now we have to prove the inductive step, i.e., that $v_{k+3} = v'_{k+1} w$. From the bi-ideal construction and inductive hypotheses $\forall i \leq k$: $v_{i+2} = v'_i w$ we get that

$$v_{k+3} = v_{k+2} u_{k+3} v_{k+2} = v'_i w u_{k+3} v'_i w.$$

Let's consider 2 cases, based on the parity of indexes:

(i) If k - even number, then

$$v'_k w u_{k+3} v'_k w = v'_k w 1 v'_k w = v'_k 101 v'_k w.$$

If we define $u'_{k+1} = 101$, we get that

$$v'_k 101 v'_k w = v'_k u'_{k+1} v'_k w = v'_{k+1} w.$$

(ii) If k - odd number, then

$$v'_k w u_{k+3} v'_k w = v'_k w 0 v'_k w = v'_k 100 v'_k w.$$

If we define $u'_{k+1} = 100$ in this case, againgwe get that

$$v'_k 100 v'_k w = v'_k u'_{k+1} v'_k w = v'_{k+1} w.$$

And that completes the proof. It means that if we choose $u'_0 = 01000$, then there exists an almost finitely generated basis – $\langle 01000, 101, 100, 101, 100, \ldots, 101, 100, \ldots \rangle$, which generates the same bi-ideal x.

Suppose that we have a given bi-ideal over a finite alphabet A and $|A|$=k. Is it possible to find a basis sequence, which corresponds to it? In general case, it is not possible and we need to know at least two things to do that.

Proposition 37. *The amount of base words has to be certain in order to find the corresponding basis.*

Proof. Let us deal with such a basis sequence:

$$\langle 0, 0, \ldots, 0, 1 \rangle. \tag{5.22}$$

If we don't know how many base words we have got, we can't know how far we have to examine the given bi-ideal. □

Proposition 38. *The maximal length of base words has to be certain in order to find the corresponding basis.*

Proof. Let us deal with such a basis sequence:

$$\langle 00000 \ldots 001 \rangle. \tag{5.23}$$

Almost the same idea applies here. If we don't know how long the longest base word is, we can't know how far we have to examine the given bi-ideal. □

Theorem 39. *It is possible to find a base for a given finitely generated bi-ideal.*

Proof. This means that we have to know two numbers:

$$n \text{ – amount of base words,}$$
$$l = \max_{0 \leq i \leq n-1} u_i.$$

From this we get that it is possible to find a base sequence only for finitely generated bi-ideals. As we know the cardinality of the alphabet, the maximum length of the base words (bounded bi-ideals) and amount of base words (finitely generated bi-ideals), there is a possibility of only finite amount of distinct bases. That means we can construct a finite amount of bi-ideals from all these distinct bases. We do that together for all distinct bases continuously until almost all of bases (except the correct one) do not match at some point to the given bi-ideal sequence. At the time when penultimate base fails, we have left with the base we were looking for. This base is the one that corresponds to the given bi-ideal. □

5.3 Filling of holes in a finitely generated bi-ideal

Suppose a finitely generated bi-ideal x is given over a finite alphabet A and there is a place (index i, which correspond to $x[i]$) with a hole in it. It means we have lost an information about what letter has to be there. Natural question appears – is it possible to find a letter which corresponds to this hole?

Definition 8. If u is a word, then $\overrightarrow{u}$ defines the same word without a first letter.

Example 11. If u=01110, then $\overrightarrow{u}$=1110.

Theorem 40. *Suppose we have a finitely generated bi-ideal x with a basis sequence (u_n). If we define*

a) $u'_0 = \overrightarrow{u_0}$
$\forall i$ *>0 :* $u'_i = u_i u_0[0]$.
Then x' is a bi-ideal genereted by basis (u'_n) besides $x' = \overrightarrow{x}$.

b) $u''_0 = u'_0 u'_1$
$\forall i$ *>0 :* $u''_i = u'_0 u'_{i+1}$.

Then basis (u''_n) defines a bi-ideal $x'' = \overrightarrow{x}$.

Proof. At first, we will prove the first part. To do that, we have to show that

$$\forall n \in \mathbb{N}\colon v'_n = \overrightarrow{v_n}$$

The base case ($k = 0$) executes clearly, because $v'_0 = u'_i = \overrightarrow{u_0} = \overrightarrow{v_0}$. Now we suppose that our statement holds for some natural k, i.e., $v'_k = \overrightarrow{v_k}$. The inductive step follows from

$$v'_{k+1} = v'_k u'_{k+1} v'_k = \overrightarrow{v_k} u'_{k+1} \overrightarrow{v_k} = \overrightarrow{v_k} u_{k+1} u_0[0] \overrightarrow{v_k} = \overrightarrow{v_k} u_{k+1} v_k = \overrightarrow{v_{k+1}}.$$

The second part directly follows from Proposition 6.

□

Now we state a theorem that helps us to prove our main result in this chapter.

Theorem 41. *If x is a finitely generated bi-ideal with a basis $\langle u_0, u_1, \ldots, u_n\rangle$, then $\overrightarrow{x}$ is also a finitely generated bi-ideal.*

Proof. We can construct a basis for a finitely generated bi-ideal $\overrightarrow{x}$ in the following way. First we define a sequence:

$$u'_0 = \overrightarrow{u_0}$$
$$u'_1 = u_1 u_0[0]$$
$$u'_2 = u_2 u_0[0]$$
$$\dots$$
$$u'_n = u_n u_0[0]$$
$$u'_{n+1} = u_0 u_0[0]$$
$$u'_{n+2} = u_1 u_0[0]$$
$$u'_{n+3} = u_2 u_0[0]$$
$$\dots$$
$$u'_{2n+1} = u_n u_0[0]$$
$$u'_{2n+2} = u_0 u_0[0]$$
$$\dots$$

In other words we can define the basis as

$$u'_0 = \overrightarrow{u_0}$$
$$\forall m \geq 1 : u'_m = u_{m \bmod (n+1)} u_0[0].$$

Now suppose that the basis $\langle u'_0, u'_1, \dots, u'_n, u'_{n+1}, \dots \rangle$ defines a bi-ideal x'. In this case this basis defines an almost finitely generated bi-ideal. It means that our base sequence is ultimately periodic. The first base word is u'_0 and then we have a periodic sequence $\langle u'_1, \dots, u'_n, u'_{n+1} \rangle$ that generates our bi-ideal x'. Secondly, we define another sequence:

$$u''_0 = u'_0 u'_1$$
$$u''_1 = u'_0 u'_2$$
$$u''_2 = u'_0 u'_3$$
$$\dots$$
$$u''_n = u'_0 u'_{n+1}$$

The basis $\langle u''_0, u''_1, \dots, u''_n \rangle$ defines a bi-ideal $x'' = \overrightarrow{x}$. To prove that, at first we prove that $x' = \overrightarrow{x}$ by mathematical induction. To do that we need to prove that

$$\forall i \colon v'_i = \overrightarrow{v_i}.$$

The base case ($k = 0$) executes clearly, because $v'_0 = u'_i = \overrightarrow{u_0} = \overrightarrow{v_0}$. Now we suppose that our statement holds for some natural k, i.e., $v'_k = \overrightarrow{v_k}$. From the defined basis sequence $\langle u'_0, u'_1, \dots, u'_n, u'_{n+1}, u'_1, \dots \rangle$ and construction of bi-ideals we get that

$$v'_{k+1} = v'_k u'_{k+1} v'_k = \overrightarrow{v_k} u'_{k+1} u_{(k+1) \bmod (n+1)} \overrightarrow{v_k} = \overrightarrow{v_k} u_{m \bmod (n+1)} u_0[0] \overrightarrow{v_k} = \overrightarrow{v_k} u_{m \bmod (n+1)} v_k =$$
$$= \overrightarrow{v_{k+1}}.$$

The last part that the basis $\langle u''_0, u''_1, \dots, u''_n \rangle$ defines a bi-ideal $x'' = \overrightarrow{x}$ again directly follows from Proposition7. □

(Cers, 2010) solved the decision problem: given two bases, decide whether they generate the same finitely generated bi-ideal.

Definition 9. We say a basis $\langle u_0, u_1, \dots, u_{m-1} \rangle$ of a finitely generated bi-ideal x is reducible if it can be changed by an application of any of the following reductions:

1. There is a word u and non-negative integers k_i, such that $u_i = u^{k_i}$ for all $i \in \overline{0, m-1}$. Then the 1-tuple $\langle u \rangle$ is also a basis of x.

2. There is a $T < m$ such that $m = k \cdot T$ for some $k \in \mathbb{N}$ and $u_i = u_{i+T}$ for all $i \in \overline{1, m-T-1}$. Then the T-tuple $\langle u_0, u_1, \dots, u_{T-1} \rangle$ is also a basis of x.

3. There are words w_i such that $u_i = w_{m-1} w_i$ for all $i \in \overline{0, m-1}$. Then the m-tuple $\langle w_{m-1}, w_0, w_1, \dots, w_{m-2} \rangle$ is also a basis of x.

Example 12. Here we give an example of each reduction:

1. If $\langle 0101, 01010101, 01 \rangle$ is a basis of x, then $\langle 01 \rangle$ also is a basis of x;

2. If $\langle 0, 1, 2, 0, 1, 2 \rangle$ is a basis of x, then $\langle 0, 1, 2 \rangle$ also is a basis of x;

3. If $\langle 01020100, 01020101, 01020102 \rangle$ is a basis of x, then $\langle 0102, 0100, 0101 \rangle$ is also a basis of x.

Definition 10. A finite basis of a bi-ideal is called *irreducible* if it cannot be further reduced by any reduction of Definition 9.

Theorem 42. *There is one and only one irreducible basis for any finitely generated bi-ideal.*

Theorem 43. *If two irreducible finitely generated bi-ideals x and y are not equal, then there*

$$\overset{\infty}{\exists} i : x_i \neq y_i.$$

.

Proof. Suppose that we have two bi-ideals – x with a basis $\langle u_0, u_1, \ldots, u_n\rangle$ and y with a basis $\langle w_0, w_1, \ldots, w_m\rangle$, with finitely many different occurrences. It means that we can find an occurrence starting from which both bi-ideals are equal. We also suppose that both of bases $\langle u_0, u_1, \ldots, u_n\rangle$ and $\langle w_0, w_1, \ldots, w_m\rangle$ are irreducible.

First, suppose that incorrespondence is only in the first letter. This means that $x = 0z$ and $y = 1z$. By the help of construction from Theorem 41, it is possible to construct two finitely generated bases for bi-ideal z. The first one:

$$\begin{aligned} u''_0 &= u'_0u'_1 = \overrightarrow{u_0}u_1u_0[0] \\ u''_1 &= u'_0u'_2 = \overrightarrow{u_0}u_2u_0[0] \\ u''_2 &= u'_0u'_3 = \overrightarrow{u_0}u_3u_0[0] \\ &\ldots \\ u''_n &= u'_0u'_{n+1} = \overrightarrow{u_0}u_{n+1}u_0[0] \end{aligned}$$

And the second one:

$$\begin{aligned} w''_0 &= w'_0w'_1 = \overrightarrow{w_0}w_1w_0[0] \\ w''_1 &= w'_0w'_2 = \overrightarrow{w_0}w_2w_0[0] \\ w''_2 &= w'_0w'_3 = \overrightarrow{w_0}w_3w_0[0] \\ &\ldots \\ w''_m &= w'_0w'_{m+1} = \overrightarrow{w_0}w_{m+1}w_0[0] \end{aligned}$$

At this point we have two bases

$$\langle u''_0, u''_1, \ldots, u''_n\rangle \text{ and } \langle w''_0, w''_1, \ldots, w''_m\rangle,$$

which generate the same finitely generated bi-ideal z. From Theorem 42 it is known that for any finitely generated bi-ideal there exists one and only one reduced basis. It means that it should be possible to reduce $\langle u''_0, u''_1, \ldots, u''_n\rangle$ to $\langle w''_0, w''_1, \ldots, w''_m\rangle$ or vice versa.

From the construction of base words $\langle u''_0, u''_1, \ldots, u''_n\rangle$, it is known that the last letter of all these base words is 0 because $u_0[0] = 0$. Likewise, it is known that the last letter of all the base words $\langle w''_0, w''_1, \ldots, w''_m\rangle$ is 1 because $w_0[0] = 1$. From the definition of reducibility of finitely generated bi-ideals, it is easy to see that in all three cases the last letter of the base words remains the same. Thus, it mean it is not possible to reduce these two bases into one. Contradiction. It is not possible that incorrespondence of two finitely generated bi-ideals is only in the first letter.

Now suppose that we have a general case – incorrespondence is in finitely many different occurrences. Let us suppose that an occurrence, starting from which both bi-ideals are equals, is some integer k+1. This mean that we can make a construction from Theorem 41 (making basis for a bi-ideal $\overrightarrow{x}$) k-times for both bi-ideals x and y and get to the case we already dealt with – $0z$ and $1z$. Doing this change of bases we still have a finitely generated bi-ideal. This concludes the proof. □

Theorem 44. *It is possible to fill the finite number of holes for a given finitely generated bi-ideal.*

Proof. As we have a finite alphabet, there is a finite number of ways to fill the holes in. It means that we have a finite amount of potential finitely generated bi-ideals. And there is still a possibility of only a finite number of distinct bases. That means we can construct a finite amount of bi-ideals from all these distinct bases. We do that for all distinct bases continuously until almost all of bases (except the correct one) do not match all the potential bi-ideal sequences. At some point there will be just one basis and one potential bi-ideal without fail. This basis is the corresponding one to the given bi-ideal and we can fill in the holes. □

We will give an example, which shows that this problem is not resolvable in general case with infinite amount of holes in a finitely generated bi-ideal.

Example 13. Suppose we have a finitely generated bi-ideal with the basis $\{00, 01\}$. This basis generates a bi-ideal

$$x = 00010000000100\ldots$$

A letter 0 is always in every odd position of this bi-ideal. Suppose that we have the holes of the given bi-ideal in every odd position

$$x = \diamond 0 \diamond 1 \diamond 0 \diamond 0 \diamond 0 \diamond 0 \diamond 1 \diamond 0 \ldots$$

It is impossible to fill these holes in because there is a finitely generated bi-ideal with a basis $\langle 10, 11\rangle$, which corresponds to the given bi-ideal.

Corollary 45. *If a given finitely generated bi-ideal x has an infinite number of holes, it is not possible to fill them all in general case.*

6 Fuzzy metrics on the set of infinite words and fuzzifying topologies

Some researchers working on the theory of automatic sequences, stringology, in particular, became interested in the use of different analytical methods to study the structure of sets of infinite words and languages. In particular, different metrics describing distance between infinite words, limits of sequences of words and topologies, both metrizable and non-metrizable, on the set of infinite words were studied. In this chapter we develop an alternative approach to the study of the analytic structure of the family of infinite words, based on the use of fuzzy metrics.

6.1 Measure of some classes of infinite words

Before we turn to the main statements and results of this chapter, let us introduce the measure of the set of infinite words and give some measures of the classes of infinite words. In Section1.2 we already mentioned the hierarchy of infinite words (see Figure 1.1). As at this point we are interested in the measure, it is defined in the following way. Let Σ be a finite alphabet. We can specify a natural topology on Σ^ω, the set of infinite words over Σ, by specifying a sub-base D as follows:

$$D = \bigcup_{\substack{j \geq 0 \\ a \in \Sigma}} D_{j,a},$$

where $D_{j,a}$ consists of those words w such that $w[j] = a$. Base elements, which are non-empty finite intersections of the $D_{j,a}$ are of the form $\Sigma^{i_1} a_1 \Sigma^{i_2} a_2 \ldots \Sigma^{i_j} a_j \Sigma^\omega$, where $j, i_1, i_2, \ldots, i_j \geq 0$ are integers and $a_1, a_2, \ldots, a_j \in \Sigma$. Such a set is called a $cylinder$. We can put a measure m on Σ^ω, by defining the measure of the cylinders:

$$m(\Sigma^{i_1}a_1\Sigma^{i_2}a_2\ldots\Sigma^{i_j}a_j\Sigma^{\omega}) = k^{-j},$$

where $k = |\Sigma|$.

In this section we are using a well–known theorem to prove our results.

Theorem 46. *Almost all sequences w over a finite alphabet Σ satisfy $p_w(n) = |\Sigma|^n$ for all $n \geq 0$.*

It means, that almost all the words w over alphabet Σ have all the possible factors of length n and as it is described above, the measure of such words is 1. From the definition of measure it follows the measure of all the infinite words is 1, because

$$m(\Sigma^{\omega}) = k^{-j} = k^0 = 1.$$

Theorem 47. *The measure of bounded bi-ideals over alphabet $\{0, 1\}$ in space of infinite words is 0.*

Proof. From the definition of bounded bi-ideals it is known that for an arbitrary bounded bi-ideal x there exists an integer l such that $\forall i\ |u_i| \leq l$. If we regard the first base word u_0 from our bi-ideal x, then there are two possible cases in an alphabet $\{0, 1\}$:
Case A
$u_0[0] = 0$, i.e., the base word u_0 starts with a letter 0. From the construction of bi-ideals is known, that

$$x = u_0u_1u_0u_2u_0u_1u_0u_3u_0u_1u_0u_2u_0u_1u_0u_4\ldots,$$

i.e., every second base word in the bi-ideal x is u_0. As it stands, that $\forall i\ |u_i| \leq l$ and $u_0[0] = 0$, then there cannot be more than 2l-1 letters of 1's between two different 0's. Hence, bi-ideal x does not contain a factor 1^{2l}. From Theorem 46 it is known, that almost all the words w over alphabet Σ satisfy $\forall n \geq 0$: $p_w(n) = |\Sigma|^n$, i.e., the measure of such words is $m(w) = 1$. In this case, our bounded bi-ideal x does not contain a factor 1^{2l}. Hence

$$m(\text{all bounded bi-ideals with the first letter } 0) = m(\Sigma^{\omega}) - 1 = 0.$$

Case B
$u_0[0] = 1$, i.e., the base word u_0 starts with a letter 1. Again from the construction of bi-ideals it is known that

$$x = u_0u_1u_0u_2u_0u_1u_0u_3u_0u_1u_0u_2u_0u_1u_0u_4\ldots,$$

and the further construction of the proof for Case B follows straightforward from Case A and therefore

$$m(\text{all bounded bi-ideals with the first letter } 1) = m(\Sigma^{\omega}) - 1 = 0.$$

□

Theorem 48. *The measure of uniformly recurrent words over alphabet* $\{0, 1\}$ *in space of infinite words is 0.*

Proof. Suppose, that x – an arbitrary uniformly recurrent word. There are 2 possible cases:

Case A

$x[0] = 0$. From the definition of uniformly recurrent words it is known that for each its factors (suppose v), there exists an integer k such that v occurs in every factor of x, which is at least of length k (see Figure 6.1). It follows, that for a factor 0 there also exists an integer $k \in \mathbb{N}$ such, that

0 occurs in x[0;k-1];
0 occurs in x[k;2k-1];
0 occurs in x[2k;3k-1];
...
0 occurs in x[nk;(n+1)k-1];
...

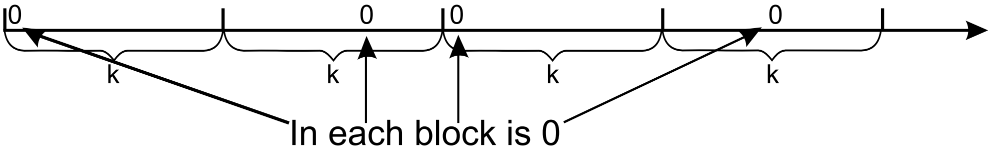

Figure 6.1: Letter 0 as a factor in a uniformly recurrent word

It means, that in every factor of length k of our uniformly recurrent word x there will appear at least one 0. Since the length of each of those blocks is k, there cannot be more than $2k$-2 letters of 1's between two consecutive 0's. It follows, that there is no factor 1^{2k-1} in our uniformly recurrent word x.

Case B

$x[0] = 1$. From the definition of uniformly recurrent words it is known that for each its factors (suppose v), there exists an integer k such that v occurs in every factor of x, which is at least of length k (see Figure 6.2). Further construction of the proof that there is no factor 0^{2k-1} in our uniformly recurrent word x follows straightforward from Case A.

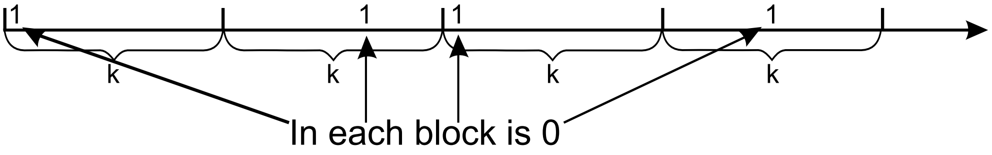

Figure 6.2: Letter 1 as a factor in a uniformly recurrent word

From Theorem 46 it is known, that almost all sequences w over a finite alphabet Σ satisfy $p_w(n) = |\Sigma|^n$ for all $n \geq 0$, i.e., the measure of such words is $m(w) = 1$. In this case, our uniformly recurrent word x does not contain a factor 1^{2k-1} in case A (a factor 0^{2k-1} in Case B). Hence,

$$m(\mathfrak{R}_{\mathrm{u}}) = m(\Sigma^\omega) - 1{=}0.$$

□

Corollary 49. *The measure of bounded bi-ideals over every finite alphabet Σ in the space of infinite words is 0.*

Corollary 50. *The measure of uniformly recurrent words over every finite alphabet Σ in the space of infinite words is 0.*

Theorem 51. *The measure of the bi-ideals over every finite alphabet Σ in the space of infinite words is 1.*

Proof. Suppose that we have a finite alphabet Σ and x – an arbitrary word with $F(x) = \Sigma^*$. From Theorem 46 it is known that the measure of all words, which have all the possible factors, is 1. So if we can show that our arbitrary word x is a bi-ideal, then we will get that measure of all bi-ideals is 1 as well. Let us take the first letter of x and denote it as u_0 and it will be the first element v_0 of our bi-ideal sequence. As the set of x factors is Σ^*, then there has to be a factor $u_0^3 = u_0u_0u_0$ somewhere in x (see Figure 6.3). Let us look at the first such one. Now we denote by u_1 the factor of x, which is between the first x letter u_0 and the last one u_0 (from first appeared factor in form $u_0u_0u_0$). From the definition of bi-ideals, we get the next element of the bi-ideal sequence $v_1 = v_0u_1v_0$, because $v_0 = u_0$.

Further, we prove theorem by induction, i.e., suppose that v_i is an element of the bi-ideal sequence. As the set of x factors is Σ^*, then there has to be a factor $v_i^3 = v_iv_iv_i$ somewhere

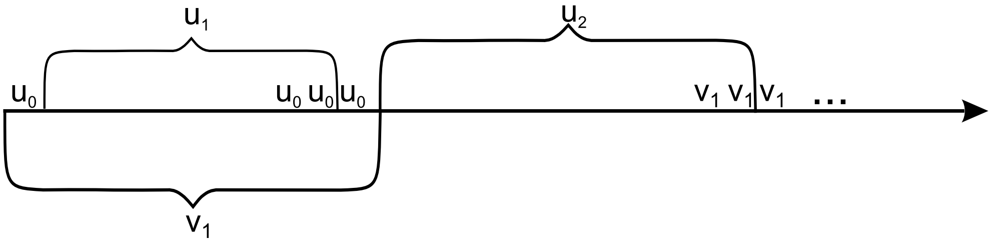

Figure 6.3: Structure of the bi-ideal x

in x. Let us look again at the first such one. Now we denote by u_{i+1} the factor of x, which is between the first x factor v_i and the last one v_i. From the definition of bi-ideals, we get the next element of bi-ideal sequence $v_{i+1} = v_i u_{i+1} v_i$. Inductive step has been done. Therefore, the arbitrary word x is a bi-ideal. Hence the measure of bi-ideals over finite alphabet is 1, i.e., $m(\mathfrak{B}) = 1$. □

6.2 Ordinary metrics on the set of infinite words

6.2.1 Pseudometrics and pseudometric space

Recall that a metric on a set X is a mapping $d : X \times X \to \mathbb{R}^+$ where $\mathbb{R}^+ = [0, \infty)$ such that for all $x, y, z \in X$:

(1d) $d(x, y) = 0 \Longleftrightarrow x = y$;

(2d) $d(x, y) = d(y, x)$;

(3d) $d(x, z) \leq d(x, y) + d(y, z)$

In case axiom (1d) is replaced by a weaker axiom

(1′d) $d(x, y) = 0 \Longleftarrow x = y$;

we come to the definition of a pseudometric.

In case a stronger version of the axiom (3d)

(3^ud) $d(x, z) \leq \max\{d(x, y), d(y, z)\}$;

holds, a pseudometric is called *an ultra pseudometric*

Clearly, every ultra pseudometric is a pseudometric, but not vice-versa: the standard metric on the plane is not an ultrametric.

A pair (X, d) where X is a set and d is a pseudometric on X is called a pseudometric space.

6.2.2 Metrics on the set of infinite words

In the literature we have found two kinds of metrics (they are ultrametrics, actually) on the set of all infinite words. The first one that we denote here by ρ, is defined as follows, see e.g. (Allouche and Shallit, 2003).

Let

$$x = (x_0, x_1, x_2, \dots x_n, \dots) \text{ and } y = (y_0, y_1, y_2, \dots y_n, \dots)$$

be infinite words. Then

$$\rho(x,y) = \begin{cases} 0 & \text{if } x = y \\ 2^{-n} & \text{otherwise, where } n = \min\{i : x_i \neq y_i\} \end{cases}$$

We do not think that this metric gives satisfactory information about real ”nearness-type” relations between the words. For example, consider the following three words $x = (1,1,1,1,1,1,...)$, $y = (0,1,1,1,1,1,1,...)$ and $z = (0,0,0,0,0,0,...)$.Then $\rho(x,y) = \rho(x,z) = 1$, that is in the both cases the distance of these infinite words is the largest possible value in the corresponding metric that equals to 1. Or, if otherwise stated, everything is dictated by the first digits of the strings. However, in different situations one's intuition may say that x should be estimated ”closer” to y than to z.

Another known definition of a metric on the set of infinite words is introduced as follows, see e.g. (Holmgren, 2000).

Let $x = (x_0, x_1, x_2, \dots x_n, \dots)$ and $y = (y_0, y_1, y_2, \dots y_n, \dots)$ be infinite words, and let for a given $i \in \mathbb{N} \cup \{0\}$ the number χ_i be defined by:

$$\chi_i(x,y) = \begin{cases} 0 & \text{if } x_i = y_i \text{ where } i \text{ is the } i\text{-th coordinate of the word} \\ 1 & \text{if } x_i \neq y_i \text{ where } i \text{ is the } i\text{-th coordinate of the word} \end{cases}$$

Now let

$$\sigma(x,y) = \sum_{i=0}^{\infty} \frac{1}{2^i} \chi_i(x,y).$$

Then one can easily see that $\sigma : X \times X \to [0,1]$ is an ultrametric on the set of all infinite words. In our opinion σ is more adequate for describing nearness of the words than ρ, since it takes into account information about the whole length of the words, not only considers the information contained in the prefixes of these words. However, this metric only gives an accumulated imformation about nearness between the words and neglects all specific details of this information. For example, let $x = (1,0,0,0,0,...)$, $y = (0,1,1,1,1,...)$ and $z = (0,0,0,0,0,...)$. Then

$\sigma(y, z) = 1$, and $\sigma(x, z) = 1$, and hence this metric does not take into account the essential difference of this words, but just accumulates all information in one number.

Therefore, we do not think that ordinary metrics is an adequate analytical tool for describing nearness-type relation between infinite words. We propose to use fuzzy metrics instead. In our opinion, which we try to justify in this thesis, fuzzy metrics are much more subtle and, if properly defined, will give more refined information about the nearness-type properties between the words.

6.3 Fuzzy metrics

In 1954 K. Menger introduced the concept of the statistical metric (Menger, 2003). The theory of statistical metric was developed mainly in the second half of the previous century by different authors, see, e.g. the fundamental monograph, (Schweizer and Sklar, 1960). Based on the concept of the statistical metric Kramosil and Michalek in (Kramosil and Michalek, 1975) introduced the notion of a fuzzy metric. Actually a fuzzy metric is in a certain sense equivalent to the concept of the statistical metric, but the essential difference is in its definition and interpretation. While the statistical metric $F_{xy}(\lambda)$ on a set X is interpreted as ”the probability that the obtained distance between points $x, y \in X$ is smaller than $\lambda \in (-\infty, +\infty)$”, the fuzzy approach to the notion of a distance follows from the idea that ”the distance between two points is not an actually existing real number, but it is *a fuzzy notion*, i.e. the only way which the distance in question is to ascribe some values from $[0, 1]$ to various sentences proclaming something related to distance” (Kramosil and Michalek, 1975).

In 1994 George and Veeramani (George and Veeramani, 1994), see also (George and Veeramani, 1997), slightly modified the original concept of a fuzzy metric, we call this modification GV-fuzzy metric. On one hand this modification allows more natural examples of fuzzy metrics, in particular fuzzy pseudometrics constructed from metrics. On the other hand George and Veeramani fuzzy pseudometrics are more appropriate for the definition and the study of the induced topological structure. In our work we modify GM-definition of a fuzzy pseudometric by weakening one of the axioms in George-Veeramani definition of a fuzzy metric thus coming to a concept which will be call a fragmentary fuzzy metric. The necessity to enlarge the class of fuzzy metrics will be explained at the appropriate place. We contruct a special fragmentary fuzzy metric on the set of infinite words from a sequence of partial ordinary pseudometrics on this set. The fuzzy metric obtained in this way will be used for the description of the analytic structure of the set of infinite words.

Before we can define the concept of a fragmentary fuzzy metric we need the notion of a t-norm, introduced first by K.Menger (Menger, 2003), and later studied and applied in the research by many authors, see e.g. (Schweizer and Sklar, 1960), (Klement et al., 2000), et. al.

6.3.1 t-norms

Definition 11. A t-norm is a binary operation on the unit interval $* : [0,1] \times [0,1] \to [0,1]$ satisfying the following conditions:

(0t) $*$ is monotone: $\alpha \leq \beta \Rightarrow \alpha * \gamma \leq \beta * \gamma$ for all $\alpha, \beta, \gamma \in [0,1]$;

(1t) $*$ is commutative: $\alpha * \beta = \beta * \alpha$ for all $\alpha, \beta \in [0,1]$;

(2t) $*$ is associative: $(\alpha * \beta) * \gamma = \alpha * (\beta * \gamma)$ for all $\alpha, \beta, \gamma \in [0,1]$;

(3t) $\alpha * 1 = \alpha, \quad \alpha * 0 = 0$ for all $\alpha \in [0,1]$;

(4t) $*$ distributes over arbitrary joins: $\alpha * \left(\bigvee_{i \in I} \beta_i\right) = \bigvee_{i \in I}(\alpha * \beta_i)$
for every $\alpha \in [0,1]$ and for all $\{\beta_i \mid i \in I\} \subseteq [0,1]$.

Example 14. Among the most important examples of t-norms are the following three:

- Let $* = \wedge$. It is called the *minimum* t-norm.
- Let $\alpha * \beta := \alpha \cdot \beta$ be the product. This is the so called *product* t-norm.
- Let $\alpha * \beta = \max(\alpha + \beta - 1, 0)$. This is the Łukasiewicz t-norm.

Remark 52. It is known that $\wedge$ is the largest t-norm:
For any t-norm $*$ and any $\alpha, \beta \in [0,1]$ it holds $\alpha * \beta \leq \alpha \wedge \beta$.

Although in order to introduce the concept of a fuzzy metric we need the general definition of the t-norm, our work is mainly based on the minimum t-norm. Referring to Remark 52 our results can be extended for the case of other t-norms if they satisfy some additional conditions.

6.3.2 Fuzzy pseudometrics and fragmentary fuzzy metrics: basic definitions and results

Definition 12. A fuzzy pseudometric on the set X is a pair $(m, *)$, or simply m, where $m : X \times X \times \mathbb{R}^+ \to [0, 1]$ (that is m is a fuzzy subset of $X \times X \times \mathbb{R}^+$), satisfying the following conditions for all $x, y, z \in X$, $s, t \in \mathbb{R}^+$:

(1GV) $m(x, y, t) > 0$;

(2GV) $m(x, y, t) = 1 \Longleftarrow x = y$;

(3GV) $m(x, y, t) = m(y, x, t)$;

(4GV) $m(x, z, t + s) \geq m(x, y, t) * m(y, z, s)$;

(5GV) $m(x, y, -) : \mathbb{R}^+ \to [0, 1]$ is continuous.

If $(m, *)$ is a fuzzy metric on X, then the triple $(X, m, *)$ is called *a fuzzy metric space*.
If axiom (2GV) is replaced by a stronger axiom

(2′GV) $x = y \Longleftrightarrow m(x, y, t) = 1$

we get definitions of a fuzzy metric, and the corresponding fuzzy metric space.

Note that axiom (4GV) combined with axiom (2GV) implies that the fuzzy metric $m(x, y, t)$ is non-decreasing on the third argument.

Definition 13. A fragmentary fuzzy metric on the set X is a pair $(m, *)$, or simply m where the mapping $m : X \times X \times \mathbb{R}^+ \to (0, 1]$ satisfies the following conditions for all $x, y, z \in X$, $s, t \in \mathbb{R}^+$:

(1FFM) $m(x, y, t) > 0$;

(2FFM) $m(x, y, t) \geq \frac{t}{t+1}$ whenever $x = y$;

(3FFM) $m(x, y, t) = m(y, x, t)$;

(4FFM) $m(x, z, t + s) \geq m(x, y, t) * m(y, z, s)$;

(5FFM) function $m(x, y, -) : \mathbb{R}^+ \to [0, 1]$ is continuous and non-decreasing.

If $(m, *)$ is a fragmentary fuzzy metric on X, then the triple $(X, m, *)$ is called *a fragmentary fuzzy metric space*.

Remark 53. Thus axioms (1FFM), (3FFM) and (4FFM) coincides with axioms (1GV), (3GV) and (4GV) respectively.

Since we weaken the axiom (2GV) to the axiom (2FFM) we had to strengthen axiom (5GV) replacing it by axiom (5FFM): The reason for this is that combination of axioms (2FFM) and (4FFM) (as different from the combination of axioms (2GV) and (4GV)) does not imply that the function $m(x, y, -) : \mathbb{R}^+ \to [0, 1]$ is non-decreasing. Therefore we have to request this important property explicitly by replacing axiom (5GV) with axiom (5FFM).

Remark 54. We think it is reasonable to replace axiom (2GV) by a weaker axiom (2FFM) at least for two reasons.

First, this generalized version of the definition of a fuzzy pseudometric is more appropriate for the description of the distance between two infinite words which is defined inductively from certain *fragments*. And second, constituting that a distance between two equal objects should be fixed for every $t \in \mathbb{R}^+$ and not to be a subject of some possible evaluation seems to be not very natural in the context of defining ”distance” with fuzzy metrics. Note also that

$$\lim_{t \to \infty} m(x, y, t) = 1 \text{ whenever } x = y$$

also in case the fragmentary fuzzy metric.

Patterned after (Piera, 2001) we introduce the following fundamental for our research concept:

Definition 14. A fragmentary fuzzy metric is called a fragmentary fuzzy ultrametric if for every $x, y, z \in X, t \in \mathbb{R}^+$:

$$m(x, y, t) \geq \min\{m(x, z, t), m(z, y, t)\}.$$

Further, the next definition is ”the fragmentary version” of the concept of a strong fuzzy metric, see e.g. (Gregori et al., 2010):

Definition 15. A fragmentary fuzzy metric m on X is called strong if, in addition to the properties (1FFM), (2FFM), (3FFM) (4FFM) and (5FFM), the following modification of axiom (4FFM) is satisfied

(4^{s}FFM) $m(x, z, t) \geq m(x, y, t) * m(y, z, t)$ for all $x, y, z \in X$ and for all $t > 0$.

To justify this definition we show that actually in this context the axiom (4FFM) may be omitted, that is axiom (4FF^{s}M) is indeed stronger than axiom (4FFM). This is proved in the next proposition:

Proposition 55. *Let* $m : X \times X \times \mathbb{R}^+$ *satisfy axioms (1FFM), (2FFM), (3FFM), (4^{s}FFM) and (5FFM). Then* $m : X \times X \times \mathbb{R}^+ \to [0, 1]$ *is a fragmentary fuzzy metric.*

Proof. Referring to axioms (4^{s}FFM) and (5FFM) we get the following series of inequalities:

$$m(x, z, t + s) \geq m(x, y, t + s) * m(y, z, t + s) \geq m(x, y, t) * m(y, z, s),$$

which holds for any $x, y, z \in X$ and any $t, s \in \mathbb{R}^+$.

□

Thus we come to the following fundamental in our research concept:

Definition 16. A strong fuzzy fragmentary metric on a set X is a pair $(m, *)$, or simply m where $m : X \times X \times \mathbb{R}^+ \to (0, 1]$, satisfies the following conditions for all $x, y, z \in X$, $t \in \mathbb{R}^+$:

(1FFM) $m(x, y, t) > 0$;

(2FFM) $m(x, y, t) \geq \frac{t}{t+1}$ whenever $x = y$;

(3FFM) $m(x, y, t) = m(y, x, t)$;

(4^{s}FFM) $m(x, z, t) \geq m(x, y, t) * m(y, z, t)$;

(5FFM) $m(x, y, -) : \mathbb{R}^+ \to [0, 1]$ is continuous and non-decreasing.

In what follows we will need the following Lemma, showing that also in this weaker form axiom (2FFM) in case of the minimum t-norm the point x is ”closer” to itself than to any other point:

Lemma 56. *Let* $(m, \wedge)$ *be a fragmentary fuzzy metric. For every* $x, y \in X$*, and every* $t \in \mathbb{R}^+$ *it holds* $m(x, x, t) \geq m(x, y, t)$*.*

Proof. From axioms (4FFM) and (3FFM) we have

$$m(x, x, t) \geq m(x, y, t) \wedge m(y, x, t) = m(x, y, t) \wedge m(x, y, t) = m(x, y, t).$$

□

Definition 17. A fragmentary fuzzy strong metric $m : X \times X \times \mathbb{R}^+ \to (0,1]$ is called a fragmentary fuzzy strong ultrametric if

$$m(x,y,t) \leq m(x,z,t) \wedge m(z,y,t)$$

for all $x, y, z \in X$, and every $t \in \mathbb{R}^+$.

Definition 18. (Gregori et al., 2010) Given two fuzzy metric spaces $(X, m, *_m)$ and $(Y, n, *_n)$, a mapping $f : X \to Y$ is called continuous if for every $\varepsilon \in (0,1)$, every $x \in X$ and every $t \in \mathbb{R}^+$ there exists $\delta \in (0,1)$ and $s \in \mathbb{R}^+$ such that $n(f(x), f(y), t) > 1 - \varepsilon$ whenever $m(x,y,s) > 1 - \delta$. In symbols:

$$\forall \varepsilon \in (0,1), \forall x \in X, \forall t \in \mathbb{R}^+ \; \exists \, \delta \in (0,1), \exists s \in \mathbb{R}^+ \text{ such that}$$

$$m(x,y,s) > 1 - \delta \Longrightarrow n(f(x), f(y), t) > 1 - \varepsilon$$

The following proposition gives the standard construction of a fuzzy metric from a usual metric on the same set:

Proposition 57. (Gregori et al., 2010) *Let (X,d) be a pseudometric space. Let m_d be the fuzzy set defined on $X \times X \times \mathbb{R}^+$ by*

$$m_d(x,y,t) = \frac{t}{t + d(x,y)}.$$

Then $(m_d, *)$ is a strong fuzzy pseudometric in case $* = \cdot$ is the product t-norm.

We will need the following modification of the above statement.

Proposition 58. *Let (X,d) be an ultrametric space and define the fuzzy set m_d on the set $X \times X \times \mathbb{R}^+ \to (0,1]$ by*

$$m_d(x,y,t) = \frac{t}{t + 1 + d(x,y)}.$$

Then $m_d(x,y,t)$ is a fragmentary strong fuzzy metric in case of t-norms $\wedge$ (minimum) and $\cdot$ (product). In particular in case of the minimum t-norm, m_d is a fragmentary fuzzy strong ultrametric.

Proof. It is clear that m_d satisfies axioms (1FFM), (2FFM) (since $d : X \times X \to [0,\infty)$ is a metric) and (3FFM). The continuity of m_d is clear and the non-decreasingness of m_d can be proved straightforward. To show (4^sFFM) let $x, y, z \in X$. Then, referring to the properties of an ultrametric $d : X \times X \to \mathbb{R}^+$ we have

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

Further, let $t > 0$ be fixed. Now we consider two separately the cases of the minimum and the product t-norms. In case $* = \wedge$ we obviously have

$$\frac{t}{t+1+d(x,z)} \geq \frac{t}{t+1+d(x,y)} \bigwedge \frac{t}{t+1+d(y,z)},$$

and hence $m_d(x,z) \geq m_d(x,y) \wedge m_d(y,z)$ and hence axiom (4FFM) holds.

In case of the product we again refer to the inequality $d(x,z) \leq \max\{d(x,y), d(y,z)\}$ and easily verify that

$$\frac{t}{t+1+d(x,z)} \geq \frac{t}{t+1+d(x,y)} \cdot \frac{t}{t+1+d(y,z)}.$$

□

Referring to Proposition 52 and to Proposition 58 we get the following

Corollary 59. *Given a ultra pseudometric $d : X \times X \to \mathbb{R}^+$ the mapping $m_d(x,y,t) = \frac{t}{t+1+d(x,y)}$ is a fuzzy fragmentary metric with respect to any t-norm $*$. In particular, if $* = \cdot$ or if $* = \wedge$, then this fuzzy fragmentary metric is strong.*

In (Gregori and Romaguera, 2004) a fuzzy pseudometric m on X is called *stationary*, if m does not depend on t, i.e. if for every $x, y \in X$, the function $m_{x,y}(t) = m(x,y,t)$ is constant. We will need the following specification of this property.

Definition 19. A fragmentary fuzzy metric m on X is said to be *stationary on the interval* $[c,d] \subseteq \mathbb{R}^+$, if for each $x, y \in X$, the function $m_{x,y}(t) = m(x,y,t)$ is constant on $[c,d]$.

6.3.3 Topology induced by a fragmentary fuzzy metrics

Let $m : X \times X \to \mathbb{R}^+ \to (0,1]$ be a fragmetary fuzzy metric. We follow the lines of the construction of a topology from a fuzzy metric, see (George and Veeramani, 1994) to define the topology induced by a fragmentary fuzzy metric.

Given a point $x \in X$, $\varepsilon \in [0,1)$, $t \in \mathbb{R}^+$ we define the ball with center x, at the level t and radius ε as follows:

$$B(x,\varepsilon,t) = \{y \mid m(x,y,t) \geq 1-\varepsilon\}$$

$$t \leq s \Longrightarrow B(x,\varepsilon,t) \subseteq B(x,\varepsilon,s) \text{ and } \varepsilon \leq \delta \Longrightarrow B(x,\varepsilon,t) \subseteq B(x,\delta,t).$$

We use the family of balls

$$\mathbb{B} = \{B(x,\varepsilon,t) \mid x \in X, \varepsilon \in [0,1), t \in \mathbb{R}^+\}$$

to induce a topology T_m^{φ}.

In (George and Veeramani, 1994), (George and Veeramani, 1997) it is proved that the family $\mathbb{B} = \{B(x, \varepsilon, t) \mid x \in X, t \in (0, \infty), \varepsilon \in (0, 1]\}$ satisfies necessary conditions to be a base for some topology T_m on X. We cannot prove the analogous theorem in our case, since the axiom (2GK) is weaker than the axiom (2FFM). Therefore, simulating the proof of such theorem we cannot guarantee that for every $y \in B(x, \varepsilon, t)$ there exists a ball $B(y, \delta, s)$ such that $B(y, \delta, s) \subseteq B(x, \varepsilon, t)$: the problem is that under our assumptions we cannot guarantee that this "ball" contains y, in particular, this "ball" can be empty. However we can prove an even stronger statement in the special case of the fragmentary fuzzy ultrametric.

Theorem 60. *Let $m : X \times X \times \times\mathbb{R}^+ \to (0, 1]$ be an fragmentary fuzzy ultrametric. Then for every $y \in B(x, \varepsilon, t)$ it holds $B(y, \varepsilon, t) \subseteq B(x, \varepsilon, t)$.*

Proof. To show that $B(y, \varepsilon, t) \subseteq B(x, \varepsilon, t)$ let $z \in B(y, \varepsilon, t)$. Then $m(x, y, t) > 1 - \varepsilon$ and, since m is fragmentary $m(x, x) \geq \frac{t}{t+1}$. Now, recalling that m is a fragmentary fuzzy ultrametric we conclude that $m(y, y, t) > 1 - \varepsilon$, that is $B(y, \varepsilon, t) \subseteq B(x, \varepsilon, t)$. Further, since $y \in B(x, \varepsilon, t)$ and $m(x, y)$ is fragmentary fuzzy ultrametric it follows that $m(y, y, t) = m(x, x, t) \geq \frac{t}{t+1}$, and hence $y \in B(y, \varepsilon, t)$.

□

One can easily verify the following proposition:

Proposition 61. *Given two fragmentary fuzzy metric spaces $(X, m, *_m)$ and $(Y, n, *_n)$, a mapping $f : (X, T_m) \to (Y, T_n)$ is continuous if and only if*
$\forall \varepsilon \in (0, 1), \forall x \in X, \forall t \in (0, \infty)\ \exists\, \delta \in (0, 1), \exists s \in (0, \infty)$ such that
$n(f(x), f(y), t) > 1 - \varepsilon$ whenever $m(x, y, s) > 1 - \delta$.

6.4 Fragmentary fuzzy ultrametric on the set of infinite words

6.4.1 Construction of a fragmentary fuzzy ultrametric on the set of infinite words

Let X be the set of infinite words. We define a sequence

$$\{d_n \mid n \in \mathbb{N} \bigcup \{0\}\}$$

of ultra pseudometrics on X as follows. Let $x = (x_0, x_1, x_2, \ldots), y = (y_0, y_1, y_2, \ldots) \in X$ and let $\chi_i(x, y) = 0$ if $x_i = y_i$ and $\chi_i(x, y) = 1$ if $x_i \neq y_i$. We define:

$d_0(x, y) = \chi_0(x, y);$

$d_1(x, y) = \chi_0(x, y) + \frac{\chi_1(x,y)}{2};$

$d_2(x, y) = \chi_0(x, y) + \frac{\chi_1(x,y)}{2} + \frac{\chi_2(x,y)}{2^2};$

$\ldots$

$d_n(x, y) = \sum_{i=0}^{n} \frac{\chi_i(x,y)}{2^n};$

$\ldots .$

Theorem 62. *Every d_n is an ultra pseudometric.*

Proof. Obviously every $\frac{\chi_i(x,y)}{2^i}$ is an ultra pseudometric. From here we conclude that every $d_n(x, y)$ is an ultra pseudometric by induction referring to the following easily provable Lemma:

Lemma 63. *Let $d_1, d_2 : X \times X \to \mathbb{R}^+$ be ultra pseudometrics. Assume that $d_1(x, y) \in \{0\} \cup [a, 1]$ for any $x, y \in X$ and that $d_2(x, y) \in [0, \frac{a}{2}]$. Then $d = d_1 + d_2 : X \times X \to [0, 1]$ is an ultra pseudometric.*

Basing on this sequence of ultra pseudometrics and referring to Proposition 58 we construct the sequence of fragmentary fuzzy strong ultrametrics on the set X of all infinite words:

$\mu_0(x, y, t) = \frac{t}{t+1+d_0(x,y)};$

$\mu_1(x, y, t) = \frac{t}{t+1+d_1(x,y)};$

$\mu_2(x, y, t) = \frac{t}{t+1+d_2(x,y)};$

$\ldots$

$\mu_n(x, y, t) = \frac{t}{t+1+d_n(x,y)};$

$\ldots .$

Further we define the following family of mappings: $m_0(x, y, t) = \mu_0(x, y, t);$

$m_1(x, y, t) = \mu_1(x, y, t) \vee \mu_0(x, y, 1);$

$m_2(x, y, t) = \mu_2(x, y, t) \vee \mu_1(x, y, 2);$

$\ldots$

$m_n(x, y, t) = \mu_n(x, y, t) \vee \mu_{n-1}(x, y, n);$

$\ldots .$

□

Proposition 64. *Mappings $m_n : X \times X \times \mathbb{R}^+ \to [0,1]$ are fragmentary strong fuzzy ultrametrics on the set X of infinite words.*

Proof. From Proposition 58 we know that each μ_n is a fragmentary fuzzy strong ultrametric, that is $\mu_n(x,z,t) \geq \mu_n(x,y,t) \wedge \mu_n(y,z,t)$. Since $m_n(x,y,t) = \mu_n(x,y,t) \vee a_n$ where a_n is some constant, it is clear that $m_n(x,y,t)$ is a fragmentary fuzzy ultrametric and besides $m_n(x,z,t) \geq m_n(x,y,t) \wedge m_n(y,z,t)$, that is m_n is strong.

□

Finally, we construct a mapping $\mathcal{M} : X \times X \times \mathbb{R}^+ \to [0,1]$ as follows:

$$\mathcal{M}(x,y,t) = \begin{cases} m_0(x,y,t) & \text{if } 0 < t \leq 1 \\ m_1(x,y,t) & \text{if } 1 < t \leq 2 \\ m_2(x,y,t) & \text{if } 2 < t \leq 3 \\ \quad \dots & \\ m_n(x,y,t) & \text{if } n < t \leq n+1 \\ \quad \dots . & \end{cases}$$

Proposition 65. *The mapping $\mathcal{M} : X \times X \times \mathbb{R}^+ \to [0,1]$ is a fragmentary fuzzy strong ultrametric.*

The proof is straightforward from Proposition 64.

6.4.2 Possible shapes of the fragmentary fuzzy strong ultrametric $\mathcal{M}$ in the first 3 stages

We illustrate the shape in the initial interval $(0,3]$ of the fuzzy metric $\mathcal{M}$ describing the distance between infinite words $x = (x_0, x_1, x_2...)$ and $y = (y_0, y_1, y_2...)$ in dependence of the values x_0, x_1, x_2, y_0, y_1 and y_2.

1. The case $x_0 = y_0, x_1 = y_1, x_2 = y_2$. Then

$$\mathcal{M}(x,y,t) = \frac{t}{t+1} \text{ for } t \in (0,3].$$

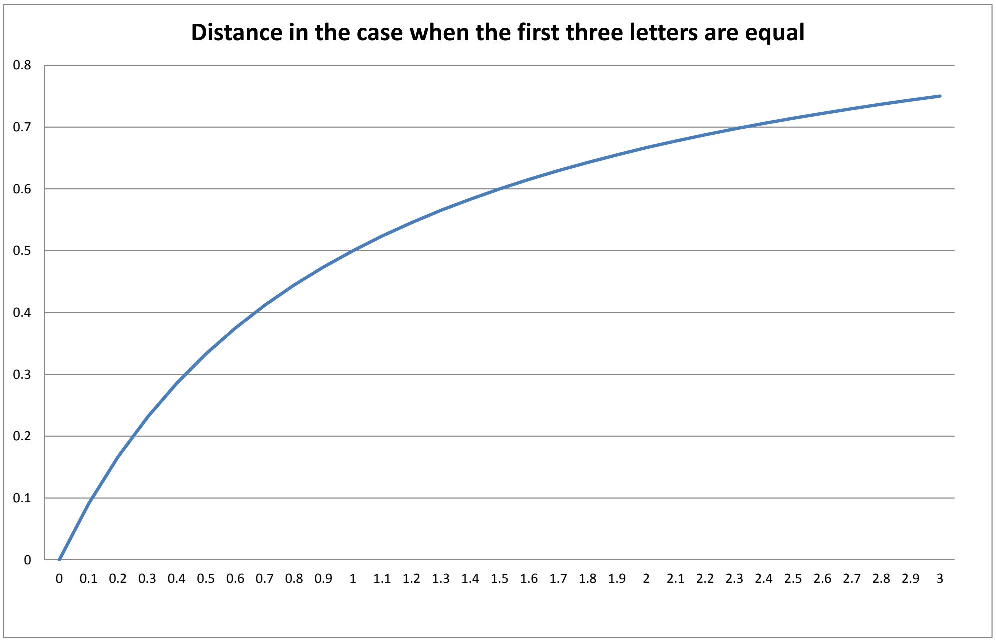

Figure 6.4: The metrics for the Case 1

2. The case $x_0 = y_0, x_1 = y_1, x_2 \neq y_2$. Then

$$\mathcal{M}(x,y,t) = \begin{cases} \frac{t}{t+1} & \text{if } 0 < t \leq 2 \\ \frac{2}{3} & \text{if } 2 < t \leq \frac{5}{2} \\ \frac{t}{t+\frac{5}{4}} & \text{if } \frac{5}{2} < t \leq 3 \end{cases}$$

3. The case $x_0 = y_0, x_1 \neq y_1, x_2 = y_2$. Then

$$\mathcal{M}(x,y,t) = \begin{cases} \frac{t}{t+1} & \text{if } 0 < t \leq 1 \\ \frac{1}{2} & \text{if } 1 < t \leq \frac{3}{2} \\ \frac{t}{t+\frac{3}{2}} & \text{if } \frac{3}{2} < t \leq 3 \end{cases}$$

4. The case $x_0 = y_0, x_1 \neq y_1, x_2 \neq y_2$. Then

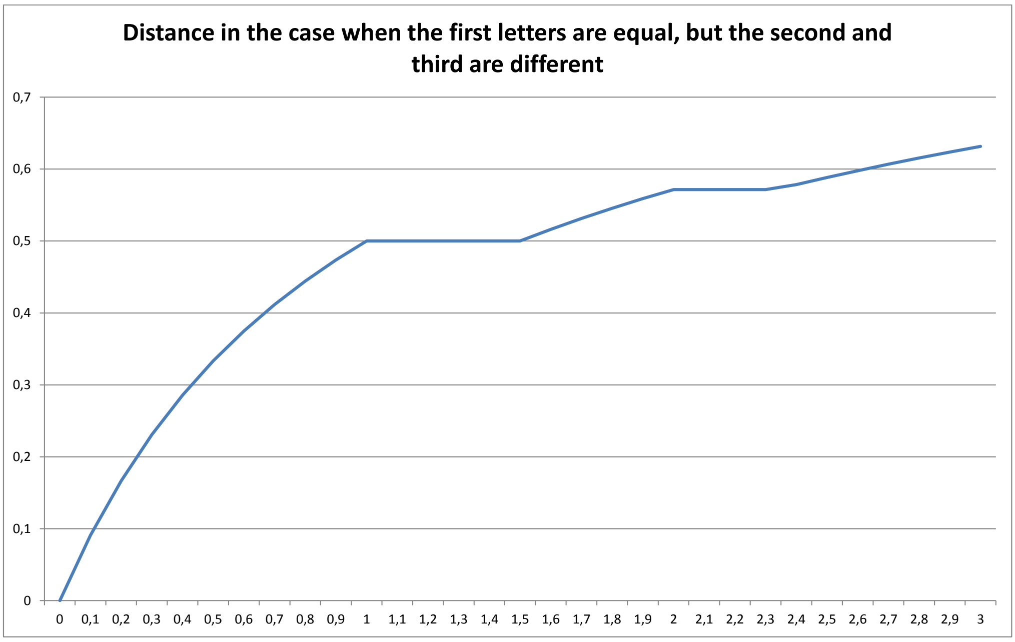

Figure 6.5: The metrics for the Case 4

$$\mathcal{M}(x,y,t) = \begin{cases} \frac{t}{t+1} & \text{if } 0 < t \leq 1 \\ \frac{1}{2} & \text{if } 1 < t \leq \frac{3}{2} \\ \frac{t}{t+\frac{3}{2}} & \text{if } \frac{3}{2} < t \leq 2 \\ \frac{4}{7} & \text{if } 2 < t \leq \frac{7}{3} \\ \frac{t}{t+\frac{7}{4}} & \text{if } \frac{7}{3} < t \leq 3 \end{cases}$$

5. The case $x_0 \neq y_0, x_1 = y_1, x_2 = y_2$. Then

$$\mathcal{M}(x,y,t) = \frac{t}{t+2} \text{ for } t \in (0,3]$$

6. The case $x_0 \neq y_0, x_1 = y_1, x_2 \neq y_2$. Then

$$\mathcal{M}(x,y,t) = \begin{cases} \frac{t}{t+2} & \text{if } 0 < t \leq 2 \\ \frac{1}{2} & \text{if } 2 < t \leq \frac{9}{4} \\ \frac{t}{t+\frac{9}{4}} & \text{if } \frac{9}{4} < t \leq 3 \end{cases}$$

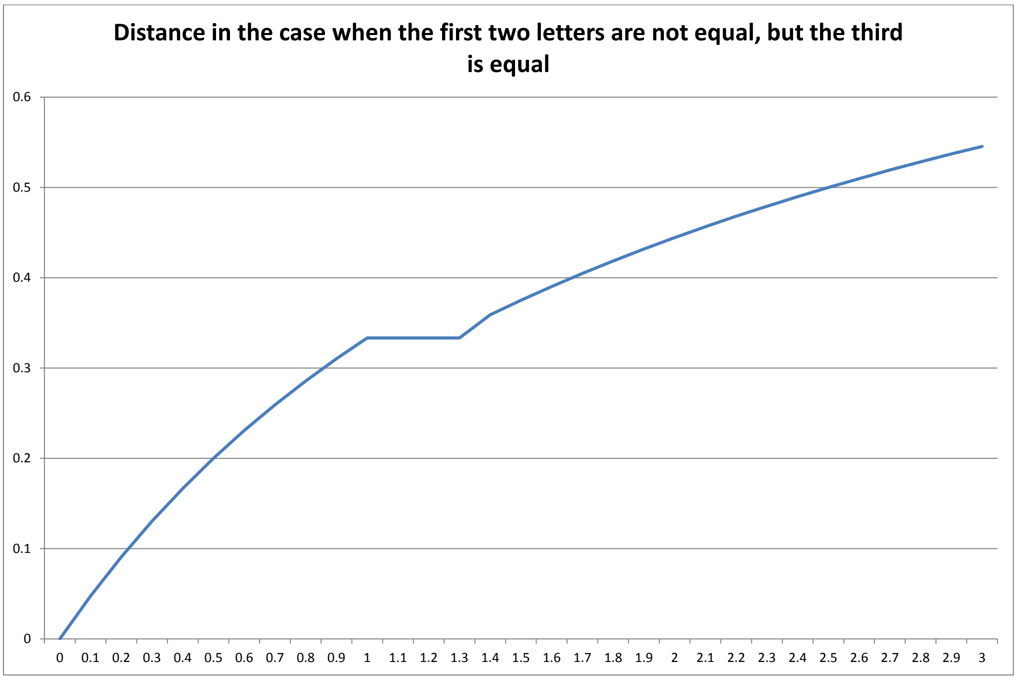

Figure 6.6: The metrics for the Case 7

7. The case $x_0 \neq y_0, x_1 \neq y_1, x_2 = y_2$. Then

$$\mathcal{M}(x,y,t) = \begin{cases} \frac{t}{t+2} & \text{if } 0 < t \leq 1 \\ \frac{1}{3} & \text{if } 1 < t \leq \frac{5}{4} \\ \frac{t}{t+\frac{5}{2}} & \text{if } \frac{5}{2} < t \leq 3 \end{cases}$$

8. The case $x_0 \neq y_0, x_1 \neq y_1, x_2 \neq y_2$. Then

$$\mathcal{M}(x,y,t) = \begin{cases} \frac{t}{t+2} & \text{if } 0 < t \leq 1 \\ \frac{1}{3} & \text{if } 1 < t \leq \frac{5}{4} \\ \frac{t}{t+\frac{5}{2}} & \text{if } \frac{5}{4} < t \leq 2 \\ \frac{4}{9} & \text{if } 2 < t \leq \frac{11}{5} \\ \frac{t}{t+\frac{11}{4}} & \text{if } \frac{11}{5} < t \leq 3 \end{cases}$$

6.5 Principal fuzzy pseudometrics

6.5.1 Topology induced by a fragmentary fuzzy strong ultrametric: the role of the property of principality

Recall that in (George and Veeramani, 1994), (George and Veeramani, 1997) a fuzzy pseudometric $(m, *)$ on a set X induces a (crisp) topology on it by taking as a base the family of all balls:

$$\mathbb{B} = \{B(x, \varepsilon, t) \mid x \in X, t \in (0, \infty), \varepsilon \in (0, 1)\}$$

where $B(x, \varepsilon, t) = \{y \in X \mid \; m(x, y, t) > 1 - \varepsilon\}$. Referring to Theorem 3.18 one can easily see that the same scheme works also in the case of a fragmentary fuzzy strong ultrametric space.

Thus, when defining a topology induced by a fragmentary fuzzy strong ultrametric space we have to take into consideration for each $x \in X$ all $\varepsilon \in (0, 1]$ as well as all $t \in \mathbb{R}^+$. The structure of the topology becomes more lucid and simple if the families $\mathbb{B}_t = \{B(x, \varepsilon, t) : x \in X, \varepsilon \in (0, 1)\}$ induce the same topology on the set X for all $t \in \mathbb{R}^+$. In other words this means that $\mathbb{B}_t = \{B(x, \varepsilon, t) : \varepsilon \in (0, 1)\}$ is a local base at the point x for the topology T_m. Following (Gregori et al., 2009) we call such fuzzy metrics *principle*. For our merits we will need the following generalization and specification of this notion.

Definition 20. A fragmentary fuzzy ultrametric $m : X \times X \times \mathbb{R}^+ \to (0, 1]$ is called principal on the interval $[c, d] \subseteq (0, \infty)$ if the families $\mathbb{B}_t = \{B(x, \varepsilon, t) : \varepsilon \in (0, 1)\}$ induce the same topology on the set X.

6.5.2 Principal fuzzy pseudometric on the set of infinite words

The following theorem will be used in the study of the topology on the set of infinite words.

Theorem 66. *Fuzzy ultrametric $\mathcal{M}$ constructed in Section 6.4 on the family of infinite words is principal.*

To prove this theorem we first establish two lemmas:

Lemma 67. *Let $d : X \times X \to [0, 1]$ be a metric and a fuzzy pseudometric $m : X \times X \times \mathbb{R}^+ \to [0, 1]$ be such that $m(x, y, t) = \frac{t}{t+1+d(x,y)}$ for each $t \in [c, d] \subseteq (0, \infty)$. Then the fuzzy pseudometric $m : X \times X \times \mathbb{R}^+ \to [0, 1]$ is principal on $[c, d]$.*

Proof. To prove this lemma it is sufficient to show that for each $x \in X$, for each $t \in [c, d]$ and for each $\varepsilon \in (0, 1)$ we can find $\delta \in (0, 1)$ such that $B(x, \delta, t) = B(x, \varepsilon, c)$. Then

$$B(x, \varepsilon, c) = \left\{ y \mid \frac{c}{c + 1 + d(x, y)} > 1 - \varepsilon \right\}$$

and

$$B(x, \delta, t) = \left\{ y \mid \frac{t}{t + 1 + d(x, y)} > 1 - \delta \right\}.$$

Since $t \in [c, d]$ we can find $\alpha \in (0, +\infty)$ such that $t = c + \alpha$. Now the requested condition that

$$B(x, \delta, t) = B(x, \varepsilon, c)$$

can be reformulated as follows: For a given α and ε we must find $\delta \in (0, 1)$ such that

$$\frac{c}{c + b} = 1 - \varepsilon \Longleftrightarrow \frac{c + \alpha}{c + \alpha + b}.$$

With simple calculations from here we get that

$$\delta = \frac{\varepsilon t}{t + \alpha(1 - \varepsilon)}.$$

Obviously $\delta = \varepsilon$ when $t = c$ and δ decreases from ε to $\frac{\varepsilon t}{t + (d - c)(1 - \varepsilon)}$ as t increases from c to d.

□

Lemma 68. *If a fragmentary fuzzy metric $m : X \times X \times \mathbb{R}^+ \to [0, 1]$ is stationary on an interval $[c, d]$, then it is also principal on this interval.*

Proof. The proof is obvious, since stationarity in this case means that $m(x, y, t) = m(x, y, s)$ for all $t, s \in [c, d]$ and hence topologies generated by all pseudometrics $m(x, y, t)$ where $t \in [c, d]$ coincide.

□

Now we are ready to prove the theorem.

Proof. From the construction of the fragmentary fuzzy ultrametric $\mathcal{M}(x, y, t)$, see Subsection 6.4.1, see also Subsection 6.4.2, it is clear that, for a given infinite word $x = (x_o, x_1, x_2, \ldots)$ this fragmentary fuzzy ultrametric defines one of the following three types of sequences of numbers

$$0 = c_0 < c_1 < c_2 < c_3 \ldots c_{2k-1} < c_{2k} < c_{2k+1} < \ldots$$

$$0 = c_0 < c_1 < c_2 < c_3 \ldots c_{2k-1} < c_{2k} < c_{2k+1}$$

$$0 = c_0 < c_1 < c_2 < c_3 \dots c_{2k-1} < c_{2k}$$

such that on the interval $(c_0, c_1]$ and on each interval $[c_{2k}, c_{2k+1}]$ for $k \in \mathbb{N} \cup \{0\}$ the fragmentary fuzzy ultrametric $\mathcal{M}(x, y, t)$ is defined by the formula $\mathcal{M}(x, y, t) = \frac{t}{t+1+d(x,y)}$ and on each interval $[c_{2k-1}, c_{2k}]$ for $k \in \mathbb{N} \cup \{0\}$ the fragmentary fuzzy ultrametric is stationary.

Consider first the case $0 = c_0 < c_1 < c_2 < c_3 \dots c_{2k-1} < c_{2k} < c_{2k+1} < \dots$. Referring to Lemma 67 we conclude that the topologies generated by the fragmentary fuzzy ultrametric $m(x, y, t)$ coincide for all $t \in (c_0, c_1]$, and all $t \in [c_{2k}, c_{2k+1}]$, $k \in \mathbb{N} \cup \{0\}$. On the other hand, referring to Lemma 68 we see that the topologies generated by the fragmentary fuzzy ultrametric $\mathcal{M}(x, y, t)$ coincide for all $t \in [c_{2k-1}, c_{2k}]$, $k \in \mathbb{N}^+$. Since the end points of the intervals belong to the both types of the intervals, by induction we conclude that the topologies generated by all $t \in (0, \infty)$ coincide and hence the fragmentary fuzzy ultrametric is principal.

In case of a finite sequence $0 = c_0 < c_1 < c_2 < c_3 \dots c_{2k-1} < c_{2k} < c_{2k+1}$ we are reasoning as in the first case and finish the proof noticing that at the last infinite interval (c_{2k+1}, ∞) the fuzzy metric is stationary.

In case of a finite sequence $0 = c_0 < c_1 < c_2 < c_3 \dots c_{2k-1} < c_{2k}$ we are reasoning as in the first case and finish the proof noticing that at the last infinite interval (c_{2k+1}, ∞) the fragmentary fuzzy ultrametric is defined by the formula $m(x, y, t) = \frac{t}{t+1+d(x,y)}$ and hence is principal.

□

6.5.3 Topology on the set of infinite words induced by the fragmentary fuzzy ultrametric $\mathcal{M}$

From Theorem 66 we immediately get the following:

Theorem 69. *For each $t \in \mathbb{R}^+$ the family $\mathbb{B} = \{B(x, \varepsilon, t) \mid \varepsilon \in (0, 1)\}$ where $B(x, \varepsilon, t) = \{y \in X \mid \frac{t}{t+1+d(x,y)} > 1 - \varepsilon\}$ is a base for the fuzzy topology $T_{\mathcal{M}}$ that is induced by the fragmentary fuzzy ultrametric $\mathcal{M}(x, y, t)$.*

Proof. Hence we can take any t in particular $t = 1$ and consider the set $\mathbb{B}_1 = \{B(x, \varepsilon, 1) : \varepsilon \in (0, 1\}$. It is easy to notice, that $\mathbb{B}_1$ contains all one-point sets $\{x\}$ of infinite words, and hence the topology induced by $\mathcal{M}$ is discrete.

□

The main conclusion from the previous theorem is the following. As different from fuzzy pseudometric $\mathcal{M}$ defined in Section 6.4, that reflects in an more or less adequate way the analytic structure of the set of all infinite words, the topology generated by this metric is discrete and hence gives us only the trivial information about the analytic structure of this set. In order to also apply topological methods for the description of the set of infinite words we suggest to use the so called fuzzifying topologies instead of ordinary topologies. This method will be developed in the next section.

6.6 Fuzzifying topologies on the set of infinite words

6.6.1 Fuzzifying topologies

The concept of a fuzzifying topology (under the name of *a fuzzy topology*) was introduced in 1980 by U. Höhle (Höhle, 1980), as a certain probabilistic modification of the concept of topology. Later, in 1991, the same concept was independently introduced by M.S. Ying (Ying, 1991), under the name of a *fuzzifying topology*. M.S. Ying rediscovered this concept by making a logical analysis of topological axioms and different properties of topological spaces. Later the theory of fuzzifying topologies got a profound development in the works by different authors, see e.g. (Ying, 1992), (Ying, 1993b), (Ying, 1993a), (Höhle, 1999), et. al.

Definition 21. Given a set X, a mapping $\mathcal{T} : 2^X \to [0,1]$ is called a fuzzifying topology on X if it satisfies the following axioms:

1. $\mathcal{T}(\emptyset) = \mathcal{T}(X) = 1$;
2. $\mathcal{T}(A \cap B) \geq \mathcal{T}(A) \wedge \mathcal{T}(B) \; \forall A, B \in 2^X$;
3. $\mathcal{T}(\bigcup_i A_i) \geq \bigwedge_i \mathcal{T}(A_i) \; \forall \{A_i : i \in I\} \subseteq 2^X$.

The pair $(X, \mathcal{T})$ is called a fuzzifying topological space.

Remark 70. The intuitive meaning of the value $\mathcal{T}(A)$ is the degree to which a set $A \subseteq X$ is open. In particular, an ordinary topology T on a set X can be realized as a fuzzifying topology $\mathcal{T} : 2^X \to \{0,1\} \subset [0,1]$ by assigning $\mathcal{T}(A) = 1$ if and only if $A \in T$, and $\mathcal{T}(A) = 0$ otherwise.

Definition 22. Given two fuzzifying topological spaces $(X, \mathcal{T}^X)$ and $(Y, \mathcal{T}^Y)$, a mapping $f : (X, \mathcal{T}^X) \to (Y, \mathcal{T}^Y)$ is called continuous if

$$\mathcal{T}^X\left(f^{-1}(B)\right) \geq \mathcal{T}^Y(B) \;\; \forall B \subseteq Y.$$

6.6.2 Fuzzifying topology on the set of infinite words

We again consider the family of fragmentary fuzzy strong ultrametrics $\{\mu_n \mid n = 0, 1, \ldots, n, \ldots\}$ (cf subsection 6.4.1): defined as follows:

$\mu_0(x, y, t) = \frac{t}{t+1+\chi_0(x,y)}$;
$\mu_1(x, y, t) = \frac{t}{t+1+\chi_1(x,y)}$;
$\mu_2(x, y, t) = \frac{t}{t+1+\chi_2(x,y)}$;
$\ldots$
$\mu_n(x, y, t) = \frac{t}{t+1+\chi_n(x,y)}$;
$\ldots.$

Further, for every $n \in \mathbb{N} \cup \{0\}$ let T_n^μ be the topology, induced by the fragmentary fuzzy ultrametric μ_n, that is T_n^μ is the topology defined by the family of balls $\mathbb{B}_n = \{B_n(x, \varepsilon, t) : x \in X, \varepsilon, t \in \mathbb{R}^+\}$ where

$$B_n(x, \varepsilon, t) = \{y \mid \mu_n(x, y, t) > 1 - \varepsilon\}.$$

Starting with the sequence of topologies $\{T_0^\mu, T_1^\mu, \ldots, T_n^\mu, \ldots\}$ we construct by induction an increasing family of topologies

$$\{T_0, T_1, \ldots, T_n, \ldots\} \text{ where } T_n = \sup\{T_0^\mu, T_1^\mu, \ldots, T_n^\mu\}$$

We extend the obtained family of topologies $\{T_0, T_1, \ldots, T_n \ldots\}$ to the family indexed by all non-negative numbers by setting $T_t = T_n \forall t \in [n, n+1)$. As the result we obtain a non-decreasing family of topologies

$$T_t : t < t', t \geq 0 \Longrightarrow T_t \subseteq T_{t'}.$$

Let $\varphi : [0, \infty) \to (0, 1]$ be any order reversing continuous bijection and let $\psi : (0, 1] \to [0, \infty)$ be its inverse (For example one can take $\varphi(t) = \frac{1}{t+1}$, then $\psi(\alpha) = \frac{1-\alpha}{\alpha}$). By setting $\tau_\alpha = T_{\psi(\alpha)}$ we obtain a non-increasing family of topologies on the set X:

$$\{\tau_\alpha : \alpha \in (0, 1]\}.$$

Theorem 71. *By setting $\mathcal{T}(A) = \sup\{\alpha \mid A \in \tau_A\}$ for each $A \subseteq X$ where X is the family of infinite words, we obtain a fuzzifying topology.*

Proof. 1. $\mathcal{T}(\emptyset) = \mathcal{T}(X) = 1$, since obviously $\emptyset, X \in T_\alpha$ for every α.

2. To show that $\mathcal{T}(A \bigcap B) \geq \mathcal{T}(A) \wedge \mathcal{T}(B)$ for any $A, B \subseteq X$ and assume that $\mathcal{T}(A) \leq \mathcal{T}(B)$. Then $\mathcal{T}(A) = \sup\{\lambda : \lambda < \mathcal{T}(A), \lambda \in (0,1]\}$, and $A, B \in T_\lambda$ for every $\lambda < \mathcal{T}(A), \lambda \in (0,1]$. Hence $A \cap B \in T_\lambda$ whenever $\lambda < \mathcal{T}(A) \wedge \mathcal{T}(B), \lambda\,(0,1]$, and therefore $(A \cap B) \geq (A) \wedge (B)$ by the definition of $\mathcal{T}$.

3. To show that $\mathcal{T}(\bigcup_i A_i) \geq \bigwedge_i \mathcal{T}(A_i)$ for every family $\{A_i : i \in I\}$, where $A_i \subseteq X$ for each $i \in I$, let $\bigwedge_i \mathcal{T}(A_i) = \alpha$. In case $\alpha = 0$ the statement is obvious. Otherwise, $\alpha = \sup\{\lambda : \lambda < \alpha, \lambda \in (0,1]\}$, and hence $\mathcal{T}(A_i) \geq \lambda$ for every $i \in I$ and every $\lambda < \alpha, \lambda \in (0,1]$. Therefore for every $\lambda < \alpha, \lambda \in K$ the family T_λ contains all A_i, $i \in I$. However, this means that $\bigcup_i A_i \in T_\lambda$ for every $\lambda < \alpha, \lambda \in (0,1]$, and hence, by the definition of $\mathcal{T}$ we have $\mathcal{T}(\bigcup_i A_i) \geq \alpha = \bigwedge_i \mathcal{T}(A_i)$.

□

Remark 72. We explain the meaning of the value $\mathcal{T}(A) \geq \alpha$ as follows. Let $\{x_{i_j} \in \{0,1\}\}$ for all $j = 1, \ldots, k$ and let $(x_{i_1} < \ldots < x_{i_k})$ let $V(x_{i_1}, \ldots, x_{i_k})$ be the family of *all words* having in the position i_j the designated value x_{j_1}. Then, given a set $A \subseteq X$ the inequality $\mathcal{T}(A) \geq \frac{1}{n+1}$ means that A can be obtained as a union of some family of sets $V(x_{i_1}, \ldots, x_{i_k})$ where $i_k \leq n$.

Conclusions

This thesis focused on investigation of subclasses of different bi-ideals. The construction of a 1-bounded bi-ideal with WELLDOC property has been given. Such infinite words are used to constuct aperiodic infinite words wit good statistical behaviour. It is important to note that it is possible to construct an infinite number of such bi-ideals.

We solved the problem of filling holes (unknown information) in finitely generated bi-ideals. We proved that not only the finite amount of holes in finitely generated bi-ideals can be recreated, but also with the help of the counterexample showed, that in the case of infinite number of holes the problem is unsolvable.

Also together with co-authors we did research on possible use of finitely generated bi-ideals in cryptography by modifying the so–called shrinking generator.

Since bi-ideals by definition are a limit of a sequence it was purposeful to investigate metric at the end of this thesis. Known metrics on infinite words poorly describe nearness-type relations, therefore a new (fuzzy) metric was introdused. The metric offered in this thesis is an area to be further developed. As currently there is a lot of research on fuzzy sets and metrics it is likely that our research could prove to be a turning point for a new broad field of research.

List of attended Conferences

International conferences

- 15th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC 2015), September 21–24, 2015, Timisoara, Romania,
  presentation – “On the Existence of 1-Bounded Bi-ideals with the WELLDOC Property”.

- 15th Central European Conference on Cryptology (CECC 2015), July 8–10, 2015, Klagenfurt, Austria,
  presentation – “WELLDOC Property in Bi-ideals”.

- 15th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC 2013), September 23–26, 2013, Timisoara, Romania,
  presentation – “Bounded Bi-ideals and Linear Recurrence”.

- 14th Mons Days of Theoretical Computer Science (JM2012), September 11–14, 2012, Louvain-La-Neuve, Belgium,
  presentation – “On the Measure of Some Classes of Infinite Words”.

- 1st International Conference of Students, Postgraduates and Young Scientists Theoretical and Applied Aspects of Cybernetics (TAAC), February 21–25, 2011, Kiev, Ukraine,
  presentation – “The Problem of Finitely Generated Bi-ideal Equality and Periodicity”.

Domestic conferences and seminars

- 73rd Conference of University of Latvia, February 19, 2015, Riga,
  presentation – “WELLDOC īpašība bi-ideālos”.

- 9th Conference of Latvian Mathematical Society, April 9–10, 2012, Jelgava,
  presentation – “Dažu bezgalīgu vārdu klašu mērs”.

- 70th Conference of University of Latvia, February 23, 2012, Riga,
  presentation – “Zelta griezums vārdu kombinatorikā”.

- 69th Conference of University of Latvia, March 17, 2011, Riga,
  presentation – “Galīgi ģenerētu bi-ideālu vienādības un periodiskuma problēma”.

Bibliography

Adjan, Sergei I. 1979. *The burnside problem and identities in groups*, Springer.

Allouche, J.-P. and J. Shallit. 2003. *Automatic sequences: Theory, applications, generalizations*, Cambridge University Press, Cambridge, UK.

Balková, Lubomíra, Michelangelo Bucci, Alessandro De Luca, and Svetlana Puzynina. 2013a. *Aperiodic pseudorandom number generators based on infinite words*, Math. Comput. **arXiv:1311.6002**.

______. 2013b. *Infinite words with well distributed occurrences*, Combinatorics on words - 9th international conference, WORDS 2013, turku, finland, september 16-20. proceedings, pp. 46–57.

Berstel, J. and J. Karhumäki. 2003. *Combinatorics on words - a tutorial.*

Berstel, J. and D. Perrin. 2007. *The origins of combinatorics on words*, European Journal of Combinatorics **28**, no. 3, 996–1022.

Blakeley, Brandon, Francine Blanchet-Sadri, Josh Gunter, and Narad Rampersad. 2009. *On the complexity of deciding avoidability of sets of partial words*, Developments in language theory, pp. 113–124 (English).

Blanchet-Sadri, F. and Ajay Chriscoe. 2004. *Local periods and binary partial words: an algorithm*, Theoretical Computer Science **314**, no. 1–2, 189 –216.

Blanchet-Sadri, F., Kevin Corcoran, and Jenell Nyberg. 2008. *Periodicity properties on partial words*, Information and Computation **206**, no. 9–10, 1057 –1064. Special Issue: 1st International Conference on Language and Automata Theory and Applications (LATA 2007).

Blanchet-Sadri, F. and R.A. Hegstrom. 2002. *Partial words and a theorem of fine and wilf revisited*, Theoretical Computer Science **270**, 401–419.

Blanchet-Sadri, F., Steven Ji, and Elizabeth Reiland. 2012. *Number of holes in unavoidable sets of partial words ii*, J. of Discrete Algorithms **14**, 65–73.

Blanchet-Sadri, F., Raphaël M. Jungers, and Justin Palumbo. 2009. *Testing avoidability on sets of partial words is hard*, Theoretical Computer Science **410**, no. 8–10, 968 –972.

Blanchet-Sadri, Francine, Aleksandar Chakarov, Lucas Manuelli, Jarett Schwartz, and Slater Stich. 2012. *Constructing partial words with subword complexities not achievable by full words.*, Theor. Comput. Sci. **432**, 21–27.

Buls, J. and A. Lorencs. 2006. *From bi-ideals to periodicity*, Proceedings of the 11th mons days of theoretical computer science, pp. 97–110.

______. 2008. *From bi-ideals to periodicity*, RAIRO-Theoretical Informatics and Applications **42**, no. 3, 467–475.

Calude, Cristian S., Helmut Jürgensen, and Ludwig Staiger. 2009. *Topology on words*, Theoretical Computer Science **410**, no. 24–25, 2323 –2335.

Cers, E. 2010. *An unique basis representation of finitely generated bi-ideals*, proceedings of the 13th mons theoretical computer science days (jm 2010), universite de picardie jules verne.

______. 2012. *Finitely generated bi-ideals and the semilattice of machine invariant ω-languages*, Ph.D. Thesis.

Chomsky, Noam and Marcel Paul Schützenberger. 1963. *The Algebraic Theory of Context-Free Languages*, Computer programming and formal systems, pp. 118–161.

Coppersmith, D., H. Krawczyk, and Y. Mansour. 1994. *The shrinking generator*, Proceedings of the 13th annual international cryptology conference on advances in cryptology, pp. 22–39.

Coudrain, M. and M.P. Schützenberger. 1966. *Une condition de finitude des monoides finiment engendres*, CR Acad. Sci., Paris, Ser. A **262**, 1149–1151.

de Luca, A. and S. Varricchio. 1999. *Finiteness and regularity in semigroups and formal languages*, Springer-Verlag, Berlin, Heidelberg.

Durand, F. 1998. *A characterization of substitutive sequences using return words*, Discrete Math. **179**, 89–101.

______. 2000. *Linearly recurrent subshifts have a finite number of non-periodic subshift factors*, Ergod. Th. and Dynam. Sys. **20**, 1061–1078.

______. 2003. *Corrigendum and addendum to: Linearly recurrent subshifts have a finite number of non-periodic subshift factors*, Ergod. Th. and Dynam. Sys. **23**, 663–669.

Durand, F., B. Host, and Skau C. 1999. *Substitution dynamical systems, bratteli diagrams and dimension groups*, Ergod. Th. and Dynam. Sys. **19**, 953–993.

Durand, F., J. Leroy, and G. Richomme. 2013. *Do the properties of an s-adic representation determine factor complexity?*, J. of Integer sequences **16**, 1–30. Article 13.2.6.

George, A. and P. Veeramani. 1994. *On some results in fuzzy metric spaces*, Fuzzy Sets and Systems **64**, no. 3, 395 –399.

______. 1997. *On some results of analysis for fuzzy metric spaces*, Fuzzy Sets and Systems **90**, no. 3, 365 –368.

Gregori, Valentín, Samuel Morillas, and Almanzor Sapena. 2010. *On a class of completable fuzzy metric spaces*, Fuzzy Sets Syst. **161**, no. 16, 2193–2205.

Gregori, Valentín, Andrés López-Crevillén, Samuel Morillas, and Almanzor Sapena. 2009. *On convergence in fuzzy metric spaces*, Topology and its Applications **156**, no. 18, 3002 –3006.

Gregori, Valentín and Salvador Romaguera. 2004. *Characterizing completable fuzzy metric spaces.*, Fuzzy Sets and Systems **144**, no. 3, 411–420.

Höhle, U. 1999. *Characterization of l-topologies by l-valued neighborhoods*, Mathematics of fuzzy sets, pp. 389–432.

Höhle, Ulrich. 1980. *Upper semicontinuous fuzzy sets and applications*, Journal of Mathematical Analysis and Applications **78**, no. 2, 659 –673.

Holmgren, A., R. 2000. *A first course in discrete dynamical systems, second edition*, Springer-Verlag.

Hu, Y., X. Liao, K.-W. Wong, and Q. Zhou. 2009. *A true random number generator based on mouse movement and chaotic cryptography*, Chaos Solitons and Fractals **40**, 2286–2293.

Karhumäki, J. 2004. *Combinatorics on words: A new challenging topic*, Technical Report 645, Turku Centre for Computer Science.

Klement, Erich Peter, Radko Mesiar, and Endre Pap. 2000. *Triangular norms*, 1st ed., Springer.

Kramosil, Ivan and Jiri Michalek. 1975. *Fuzzy metrics and statistical metric spaces.*, Kybernetika **11**, no. 5, 336–344.

L'Ecuyer, P. 1998. *Random number generation*, Handbook on simulation.

Lorencs, A. 2012. *The identity problem of finitely generated bi-ideals*, Acta Informatica **49**, no. 2, 105–115.

Lothaire, M. 1983. *Combinatorics on words. encyclopedia of mathematics and its applications 17*, Cambridge University Press, Cambridge, UK.

——— . 2002. *Algebraic combinatorics on words. encyclopedia of mathematics 90*, Cambridge University Press, Cambridge, UK.

Marsaglia, G. 1996. *Diehard: a battery of tests of randomness*, See http://www.stat.fsu.edu/pub/diehard/.

Menger, Karl. 2003. *Probabilistic geometry*, Selecta mathematica, pp. 441–444 (English).

Morse, M. and G.A. Hedlund. 1938. *Symbolic dynamics*, Amer. J.Math. **60**, 815–866.

——— . 1940. *Symbolic dynamics ii*, Amer. J.Math. **62**, 1–42.

Novikov, P.S. 1955. *On the algorithmic unsolvability of the word problem in group theory*, Tr. Mat. Inst. Steklova 55.

Oishi, S. and H. Inoue. 1982. *Pseudo-random number generators and chaos*, Transactions of the Institute of Electronics and Communication Engineers of Japan E **65**, 534–541.

Patidar, V., K. K. Sud, and N. K. Pareek. 2009. *A pseudo random bit generator based on chaotic logistic map and its statistical testing*, Informatica **33**, no. 4, 441–452.

Phatak, S.C. and S. Suresh Rao. 1995. *Logistic map: A possible random number generator*, Physical Review E **51**, no. 4, 3670–3678.

Piera, Almanzor Sapena. 2001. *A contribution to the study of fuzzy metric spaces*, Applied General Topology **2**, no. 1, 63–75.

Sandri, G.H. 1992. *A simple nonperiodic random number generator: A recursive model for the logistic map*, Technical Report GL-TR-89-1066, Boston University College of Engineering and Center for Space Physics Boston.

Schneier, B. and P. Sutherland. 1995. *Applied cryptography: protocols, algorithms, and source code in c*, John Wiley & Sons, Inc. New York, NY, USA.

Schützenberger, M.P. 1965. *On a factorization of free monoids*, Proc. Amer. Math. Soc. **16**, 21–24.

Schützenberger, M.P.l. 1956. *Une théorie algébrique du codage*, Séminaire Dubreil. Algèbre et Théorie des Nombres **9**, 1–24.

Schweizer, B. and A. Sklar. 1960. *Statistical metric spaces.*, Pacific J. Math. **10**, no. 1, 215–229.

Simon, I. 1988. *Infinite words and a theorem of hindman*, Rev. Mat. Apl. **9**, 97–104.

Thue, A. 1906. *Über unendliche zeichenreihen*, Norske Vid. Selsk. Skr. I Math-Nat. Kl. 7, 1–22.

______. 1912. *Über die gegenseitige loge gleicher teile gewisser zeichenreihen*, Norske Vid. Selsk. Skr. IMath-Nat. Kl. Chris. 1, 1–67.

Ying, Mingsheng. 1991. *A new approach for fuzzy topology (i)*, Fuzzy Sets and Systems **39**, no. 3, 303 –321.

______. 1992. *A new approach for fuzzy topology (ii)*, Fuzzy Sets and Systems **47**, no. 2, 221 –232.

______. 1993a. *Compactness in fuzzifying topology*, Fuzzy Sets and Systems **55**, no. 1, 79 –92.

______. 1993b. *A new approach for fuzzy topology (iii)*, Fuzzy Sets and Systems **55**, no. 2, 193 –207.

Zimin, A.I. 1982. *Blocking sets of terms*, Matematicheskii Sbornik **161**, no. 3, 363–375.

Author's publications

Bēts R., *The Problem of Finitely Generated Bi-ideal Equality and Periodicity*, In: proceedings of the 4-th International Scientific Conference Applied Information and Communication Technologies, 361–365, 2010. SCOPUS
Bērziņa I., Bēts R., Buls J., Cers E. and Kuleša L. *On a non-periodic shrinking generator.* In: proceedings of the 13^{th} International symposium on symbolic and numeric algorithms for scientific computing (SYNASC 2011), IEEE Computer Society, 348–354, 2011. SCOPUS

Bēts R., *The Problem of Finitely Generated Bi-ideal equality and periodicity.* In: proceedings of 1st International conference of students, postgraduates and young scientists Theoretical and Applied Aspects of Cybernetics (TAAC), p. 12–15, Kiev (Ukraine), 2011.
Bēts R., *On the Measure of Some Classes of Infinite Words*, 14th Mons Days of Theoretical Computer Science (JM2012), September 11–14, 2012, Louvain-La-Neuve, Belgium. Extended abstract (7 pp.).

Bērziņa I., Buls J., Bēts R. *Bounded Bi-ideals and Linear Recurrence.* In: proceedings of the 15^{th} International symposium on symbolic and numeric algorithms for scientific computing (SYNASC 2013), IEEE Computer Society, 285-392, 2013. SCOPUS

Buls J., Bēts R. *On Existence of 1-Bounded Bi-ideals with WELLDOC Property* In: proceedings of the 17th International symposium on symbolic and numeric algorithms for scientific computing (SYNASC 2015), accepted.