LATVIJAS UNIVERSITATE

UNIVERSITY OF LATVIA
FACULTY OF PHYSICS AND MATHEMATICS
DEPARTMENT OF MATHEMATICS

FINITELY GENERATED BI-IDEALS AND THE SEMILATTICE OF MACHINE INVARIANT ω-LANGUAGES

PhD thesis

Author: **Edmunds Cers**

Student ID MaSt990057

Scientific advisor: Asoc. prof. Jānis Buls

RIGA 2012

Anotācija

Disertācijā pētītas bezgalīgu valodu (ω-valodu) īpašības.

Galvenais pētījuma objekts ir galīgi ģenerēti biideāli — rekurentu vārdu (biideālu) apakšklase, kuru iespējams aprakstīt, lietojot periodisku bāzes virkni. Interesi par šiem vārdiem palielina dažas daudzsološas to kriptogrāfiskas īpašības, kas arī ir parādītas darbā.

Disertācijas galvenais rezultāts ir izšķiramības problēmas risinājums šiem vārdiem. Atrasts efektīvs algoritms, kas sniedz atbildi uz jautājumu, vai divas periodiskas biideālu bāzes ģenerē vienu un to pašu bezgalīgo vārdu. Darbā atspoguļoti arī vairāki citi jauni ar biideāliem saistīti rezultāti. Ir parādīts, ka galīgi ģenerēti biideāli ir morfisku vārdu apakšklase un, ka tie ir slēgti attiecībā pret pārveidojumiem ar nobīdes operatoru un morfismiem. Tiek sniegts piemērs, kas parāda, ka galīgi ģenerēti biideāli nav slēgti attiecībā pret pārveidošanu ar transformatoru. Piemērā galīgi ģenerēts biideāls ar transformatora palīdzību tiek pārveidots par Tue-Morsa vārdu.

Otra darba daļa saistīta ar mašīnu invarianto ω-vārdu klašu struktūras pētījumiem. Jau agrāk pierādīts, ka ω-valodas, kas ir slēgtas attiecībā pret pārveidojumiem ar Mīlija mašīnu veido augšējo pusrežģi ar platumu $2^{\aleph_0}$. Darbā parādu, ka šis pusrežģis nav modulārs, un kā sekas, nav distributīvs.

Atslēgas vārdi: galīgi ģenerēts biideāls, ω-valoda, Mīlija mašīna

Abstract

This thesis explores some properties of right infinite words, known also as ω-words.

The main subject of investigation are finitely generated bi-ideals — a subclass of recurrent words (bi-ideals) that can be described using a periodic basis sequence. The interest in these words is increased by some promising cryptographic properties demonstrated in the thesis.

The main result solves a decision problem about these words by giving an effective algorithm to answer the question — given two periodic bases for a bi-ideal, do these generate the same ω-word. Some more new results about bi-ideals are also presented in the thesis. It is shown that finitely generated bi-ideals are in fact a subclass of morphic words and that they are closed under right shift and transformation by morphism. An example showing that finitely generated bi-ideals are not closed under transformation by a transducer is given, transforming a finitely generated bi-ideal into the Thue-Morse word.

Additionally, the algebraic structure of ω-languages invariant under transformation by Mealy machines is explored. It is known that these classes form a join-semilattice of $2^{\aleph_0}$ width. I show that this semilattice is not modular and by implication, not distributive.

Keywords: finitely generated bi-ideal, ω-language, Mealy machine.

Contents

Introduction

The focus of this thesis are right-infinite words (infinite sequences of letters that have a beginning) — the so called ω-words — and their behaviour when they are transformed by transducers or Mealy machines — automata transforming one word into another.

There are several kinds of properties that are typically studied about a class of words (a language), be it finite or infinite. These include categorization — the study of the relationships between different languages; representation — different ways of describing the words belonging to a language; closure properties — the invariance of the language when certain transformations are applied, and others.

One property peculiar to ω-words is the rich structure of the so called machine invariant classes (Buls, 2003) — languages that are invariant under transformation by Mealy machines (Mealy, 1955) — simple automata that transform ω-words into ω-words. These classes form a join-semilattice of $2^{\aleph_0}$ width (Belovs, 2008). This structure is interesting as it bears some similarity to the Turing degrees (Kleene and Post, 1954). I show in this thesis that this semilattice is not modular (and, therefore, not distributive).

The periodic words — words consisting of a infinitely repeated sequence of characters — can be considered the trivial case of ω-words. There are several ways to produce non-trivial ω-words — a popular approach is to extend known classes of (finite) languages into infinite words. Non-deterministic Büchi and Muller automata can be understood to extend the notion of regular languages to infinite words (Perrin and Pin, 2002). For some recent research in this area see, for example, Diekert and Kufleitner (2011).

The other typical approach is to define constructive algorithms that produce the infinite words. One can obtain an infinite sequence of symbols — as is the case with Sturmian words (Morse and Hedlund, 1940) — in which case the sequence itself is concatenated into an infinite word; or one can look at infinite convergent sequences of words (in the sense that subsequent elements of the sequence prolong the previous words). Ex-

amples of sequences of words that can be extended into infinite words this way include the automatic (Büchi, 1960) and the morphic (Cobham, 1972) sequences yielding the automatic and morphic words, respectively. The main part of the thesis consists of the study of another such natural extension — the extension of bi-ideal sequences into bi-ideal words (called simply bi-ideals in this text).

Bi-ideal sequences are sequences 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 as both its prefix and suffix (see Definition 1.7) The words in a bi-ideal sequence are also known as sesquipowers (Simon, 1988) or Zimin's words (Zimin, 1982) and have been very useful in Algebra (Restivo and Reutenauer, 1984). The term bi-ideal sequence was introduced by Coudrain and Schützenberger (1966). Bi-ideal words are interesting in their own right due to a connection to recurrent words — words containing each of their factors an infinite number of times (Definition 1.5). It turns out a word is recurrent if and only if it is a bi-ideal.

Another property that makes the study of bi-ideals interesting is that any sequence of words (called a basis) can be used in a natural way to generate a bi-ideal (Definition 1.8). In this thesis I take an in-depth look at the case when the base sequence is periodic. We call bi-ideals with such bases finitely generated.

Another possible motivation for studying finitely generated bi-ideals are their possible uses in cryptography. We consider using them as a drop-in replacement for some periodic words, obtaining aperiodic pseudo-random sequences with good statistical properties. I have shown in Cers (2008) that under some restrictions of the basis, bi-ideals can be made to have the same asymptotic proportion of each sub-sequence of up to a fixed length, which is a desirable property for cryptographic applications. In the thesis some new results by Berzina et al. (2011) (of which I am a co-author) are presented, giving some restrictions of the basis such that when a normally periodic sequence used in a kind of cryptographic pseudo random number generator known as the shrinking generator (Coppersmith et al., 1994) is replaced by this kind of bi-ideal, the resultant pseudo random number sequence is aperiodic. It is demonstrated by practical testing that the sequence has statistical properties well suited for cryptographic applications.

It turns out, that even having periodic bases, the bi-ideals don't have to be simple. As is with bi-ideals in general, any finitely generated bi-ideal can have an infinite number

($\aleph_0$) of different bases. This gives rise to a natural decision problem — given two bases, decide whether they generate the same bi-ideal. The answer to this problem (an effective decision procedure) is the main result of this thesis. It rests on the work of Lorencs (2012), who found a solution for the case when the bi-ideal sequence is periodic with period two, and on the results of Buls and Lorencs (2008) who gave an effective procedure for a closely related decision problem — given a basis for a bi-ideal, decide whether the generated bi-ideal is a periodic word.

Further, several more new results concerning finitely generated bi-ideals are also presented in the thesis. We show that finitely generated bi-ideals are in fact a subclass of morphic words and that they are closed under left shift (dropping the first letter of the bi-ideal) and transformation by morphism (if the transformed word is still infinite). While it was known that the class of finitely generated bi-ideals is not closed under transformation by a transducer, a neat new example is given, showing how to transform a finitely generated bi-ideal into the famous Thue-Morse word (Morse, 1921; Allouche and Shallit, 1999).

From a slightly broader perspective, some more general properties of bi-ideals are explored as well. Buls (2005) has shown that uniformly recurrent words are closed under transformation by Mealy machines. I use some insights gained in Muchnik et al. (2003) to extend the result to the more general case of transducers.

Goals and objectives

The main goal of the thesis is to explore fundamental properties of bi-ideals and their possible uses in cryptography. Additionally, some properties of the algebraic structure of machine invariant classes is studied as well.

The tasks associated with the goals are

1. find a solution to the decision problem of two bases generating the same bi-ideal;
2. describe the class of finitely generated bi-ideals as they relate to other classes of ω-words;
3. describe the closure properties of finitely generated bi-ideals;
4. explore cryptographic and other potential applications of bi-ideals;
5. study such basic algebraic properties of the semi-lattice of machine invariant ω-words

as distributivity, modularity, and the like.

The scientific importance of the thesis

In the thesis, I successfully solve a fundamental decision problem in finitely generated bi-ideals — I give an algorithm for deciding whether two finite bases generate the same bi-ideal. The class of finitely generated bi-ideals was originally studied by Buls and Lorencs (2008) and is a natural subclass of the widely studied class of recurrent words. Several further results about finitely generated bi-ideals are also obtained, giving a better understanding of this class of words. A review of some motivational examples of potential uses of bi-ideals in cryptography is also given.

The other main result of the thesis extends our understanding of the semilattice of machine invariant words originally studied by (Buls, 2003). I show that this semilattice is not modular. While being a negative result, it is nonetheless important, because it is a fundamental algebraic property usually studied about semilattices.

The structure of the thesis

The thesis is organized as follows:

- Chapter 1 gives the preliminaries necessary for the whole thesis and also lists some of the existing results we build on. Included in this chapter is the fundamental result solving the decision problem of whether a bi-ideal is periodic by Buls and Lorencs (2008), upon which we build our main result.

- Chapter 2 deals with the possible applications of bi-ideals in cryptography, and so serves to illustrate the practical motivation for our work. Included in this chapter are the results of my masters thesis showing how bi-ideals with asymptotically equal number of ones and zeroes can be built by selecting specific bases. A more recent result (where I am a co-author) showing how bi-ideals can be used in a shrinking generator to obtain an aperiodic pseudo-random sequence is also presented.

- Chapter 3 contains my results on bi-ideals including the decision algorithm for the equivalence of two bases for finitely-generated bi-ideals. This chapter contains the main new results of the thesis.

- Chapter 4 presents the proof that the semilattice of machine invariant ω-words is not modular (and, therefore, that it is not distributive).

Approbation

The results presented in the thesis have been presented at 4 international conferences — both main results of the thesis (the bi-ideal basis equality problem and the lack of modularity of the semilattice of machine invariant words) have been presented at the 13^{th} Mons Theoretical Computer Science Days in Amiens, France (2010). An earlier result showing that the semilattice of machine invariant words is not distributive has been presented in the 79^{th} Workshop for General Algebra in Olomouc, Czech republic (2010).

Results regarding the cryptographic potential of bi-ideals have been presented at the 7^{th} Central European Conference on Cryptology in Smolenice, Slovakia (2007), and at the 13^{th} International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, held 2011 in Timisoara, Romania. In Romania, the results were presented by I.Bērziņa.

A list of publications with the results presented in this thesis is available in the bibliography.

1 Some known results

1.1 Preliminaries

Given the integers $a_1, a_2, \ldots, a_n$ let $\operatorname{lcm}(a_1, a_2, \ldots, a_n)$ and $\gcd(a_1, a_2, \ldots, a_n)$ denote the least common multiple and the greatest common divisor of the numbers $a_1, a_2, \ldots, a_n$, respectively. Given a number $x \in \mathbb{R}$, denote by $\lfloor x \rfloor$ the greatest integer less than or equal to x and by $\lceil x \rceil$ the least integer greater than or equal to x.

We call a finite set A an *alphabet*. A string of n letters $u = a_0 a_1 \ldots a_{n-1}$ from alphabet A is called a *finite word* of *length* n. We denote the set of all finite words from A by A^*. The length of a finite word u is denoted by $|u|$. A total map $x : \mathbb{N} \to A$ is called a (right infinite) *ω-word* and the set of ω-words from A is denoted by A^ω. Usually the letters u, v, w will be used to denote finite words and the letters x, y, z to denote (possibly) infinite words. The i-th letter (counting from 0) of a word x is denoted by $x[i]$.

Let $uv = a_1 a_2 \ldots a_{n-1} b_1 b_2 \ldots b_{m-1}$ denote word *concatenation* of the words $u = a_1 a_2 \ldots a_{n-1}$ and $v = b_1 b_2 \ldots b_{m-1}$. This operation extends naturally to the case when the right word is infinite. We denote the *empty word* by λ. $|\lambda| = 0$ and $\lambda u = u$ (in the case of finite words also $u\lambda = u$). We call a finite word u a *prefix* of the (finite or infinite) word x, if there is a word y such that $x = uy$. A (in-)finite word u is a suffix of a (in-)finite word v if there is a word w such that $v = wu$. The set of all prefixes of a word x is denoted by $\operatorname{Pref} x$ while the set of all suffixes is denoted by $\operatorname{Suff} x$. If a word x can be written as vuy (with v and y possibly empty, y possibly infinite), then we say that u is a *factor* of x and write $u \setminus x$. We say that a factor u *occurs* in x at position i if $x = vuy$ for some v and y with $|v| = i$. We denote a factor of x of length n occurring at position i by $x[i, i+n-1]$.

Definition 1.1. *A word $x \in A^\omega$ is* periodic *with* period p*, if $x[i] = x[i+p]$ for all positions $i \in \mathbb{N}$. A word $y = ux$ is* ultimately periodic *with period p if x is periodic with period p. A word is* aperiodic*, if it is not ultimately periodic. A finite word is (ultimately) periodic*

with period p if it is a prefix of a (ultimately) periodic infinite word with period p.

Theorem 1.2 (Lyndon-Schützenberger (1962))**.** *Let $x, y, z \in \Sigma^*$ and let x and z not be empty. Then $xy = yz$ if and only if $xy = yz$ is periodic with period $|x| = |z|$.*

Theorem 1.3 (Fine-Wilf (1965))**.** *Let w be a word having periods p and q and denote by $\gcd(p, q)$ the greatest common divisor of p and q. If $|w| \geq p + q - \gcd(p, q)$, then w has also the period $\gcd(p, q)$.*

Corollary 1.4. *Let uv and vw be periodic words with periods p and q. If $|v| \geq p + q - \gcd(p, q)$, then uvw is periodic with period $\gcd(p, q)$.*

Proof. Since $|v| \geq p+q-\gcd(p, q)$ then (Theorem 1.3) v is periodic with period $\gcd(p, q)$.

First we prove by induction on $|u|$ that uv is periodic with period $\gcd(p, q)$.

(i) Let $v = v_1 v_2 \dots v_k$. If $|u| = 1$ then $u = v_0$ for some letter v_0. Since the period of uv is p then $v_0 = v_p$.

Since the period of v is $\gcd(p, q)$ then $v_p = v_{\gcd(p,q)}$. Hence $v_0 = v_{\gcd(p,q)}$.

(ii) If $|u| > 1$ then u can be represented as the concatenation $au' = u$, where $|u'| + 1 = |u|$. By assumption, the period of $u'v$ is $\gcd(p, q)$. Now it follows from (i) that the period of $au'v$ is $\gcd(p, q)$ too. We have completed the inductive step.

The proof that vw is periodic with period $\gcd(p, q)$ proceeds analogously to the above and we can use Theorem 1.2 to conclude that uvw is periodic with period $\gcd(p, q)$.

□

Definition 1.5. *A word $x \in A^\omega$ is called* recurrent*, if each finite factor of it occurs in x at an infinite number of positions. A word uy, where $u \in A^*$ and $y \in A^\omega$, is called* almost recurrent*, if y is recurrent.*

Theorem 1.6. *If x is recurrent and ultimately periodic then x is periodic.*

Proof. Assume that $x = uv^\omega$, where v is the shortest period of v^ω. If $|u| = 0$ then x is periodic and we are done. If $|u| > 0$ then there must be an integer $k > 1$ and words μ and ν such that

$$v^k = \mu u v \nu,$$

implying that for some $\ell \geq 0$

$$uv = v_1 v^\ell v_2, \tag{1.1}$$

where v_1 is a suffix of v, while v_2 is a prefix of v. From the previous we can express $v = v_1' v_1 = v_2 v_2'$ for some v_1' and v_2'. Moreover, from (1.1) we find that

$$v = v_2 v_2' = v_2' v_2.$$

Unless v_2 is empty or $v_2 = v$, v cannot be the shortest period of v^ω giving a contradiction. If $v_2 = v$, then $u = v_1 v^\ell = v_1 (v_1' v_1)^\ell$ and $x = uv^\omega = (v_1 v_1')^\omega$. Finally, if v_2 is the empty word, then $uv = v_1 v^\ell$ and since $v = v_1' v_1$ this implies $uv^\omega = v_1 (v_1' v_1)^\omega = (v_1 v_1')^\omega$. □

The previous theorem means, that we don't have to consider ultimately periodic words when dealing with recurrent words. Any non-periodic recurrent word will be aperiodic.

Definition 1.7. *A sequence of finite words $v_0, v_1, \ldots, v_n, \ldots$ is called a bi-ideal sequence, if there is a sequence of finite words $u_0, u_1, \ldots, u_n, \ldots$ (with $u_0 \neq \lambda$) such that*

$$\begin{aligned} v_0 &= u_0, \\ v_{i+1} &= v_i u_{i+1} v_i. \end{aligned}$$

In this case the sequence $u_0, u_1, \ldots, u_n, \ldots$ is called the basis *of the bi-ideal sequence v.*

Note that if (v_i) is a bi-ideal sequence then $v_j \in \operatorname{Pref} v_i$ for all $j \leq i$ and also $|v_{i+1}| \geq 2|v_i|$ (since $|v_0| > 0$).

Definition 1.8. *Let (u_i) be a basis of a bi-ideal sequence (v_i) as per Definition 1.7. The limit of the bi-ideal sequence $x = \lim_{i \to \infty} v_i$ is an infinite word called a* bi-ideal. *We say, that (u_i) is a* basis *of the bi-ideal x or that x is the bi-ideal generated by the sequence (u_i).*

Note that there can be many bi-ideal sequences with the same limit and, therefore, many bases for the same bi-ideal. For example, consider the bi-ideal sequence (v_i) and a derived sequence (v_i') defined as $v_i' = v_{i+1}$. It is clear that the new sequence is also a bi-ideal sequence and that the limits of these sequences are the same.

Observation 1.9. *Let x be the bi-ideal generated by a basis (u_i) and let (v_i) be the associated bi-ideal sequence. Then for any given $j \in \mathbb{N}$ it is possible to factorize*

$$x = v_j \tilde{u}_1 v_j \tilde{u}_2 \cdots v_j \tilde{u}_n \cdots,$$

where $\tilde{u}_i \in \{u_0, u_1, \ldots, u_n, \ldots\}$ for all i.

Theorem 1.10. *A word is recurrent if and only if it is a bi-ideal.*

Proof. First assume x is recurrent. Then we can denote $v_0 = u_0 = x[0,0]$. Since x is recurrent, there will be a position $i_0 > 0$, such that $u_0 = x[i_0, i_0]$. Denote $u_1 = x[1, i_0 - 1]$ and $v_1 = u_0u_1u_0$. Since x is recurrent, there will be a position $i_1 > i_0$, such that $v_1 = x[i_1, i_1 + i_0 + 1]$. We will denote $u_2 = x[i_0 + 1, i_1 - 1]$ and $v_2 = v_1u_2v_1$. It is clear we can continue this construction to construct the sequences (v_i) and (u_i) and that (v_i) will be a bi-ideal sequence and (u_i) will be its base. But then $x = \lim_{i\to\infty} v_i$ and x is a bi-ideal.

For the other direction assume x is a bi-ideal and (v_i) is a bi-ideal sequence associated with it. For any factor $x[a,b]$ of x we can choose a k such that $|v_k| > b$. In this case $x[a,b]$ is also a factor of v_k. Since there are clearly infinitely many occurrences of v_k in x (from Observation 1.9), this means that $x[a,b]$ also occurs infinitely often in x and so x is recurrent. □

Definition 1.11. *We say that a bi-ideal x is* l-restricted *if there is a basis (u_i) generating x, such that $|u_i| \leq l$ for all $i \in \mathbb{N}$. We say that a bi-ideal is* restricted *if there is a $l \in \mathbb{N}$ such that x is l-restricted.*

Definition 1.12. *We say that a bi-ideal x is* finitely generated *if there is a periodic sequence (u_i) (there is a $T \geq 1$ such that $u_{i+T} = u_i$ for all $i \in \mathbb{N}$) such that (u_i) is a basis of x.*

An obvious consequence of Definitions 1.11 and 1.12 is that every finitely generated bi-ideal is restricted.

Definition 1.13. *If (u_i) is a periodic sequence with period n (that is $u_i = u_{i+n}$ for all $i \in \mathbb{N}$) that generates the finitely generated bi-ideal x then we call the tuple $(u_0, u_1, \ldots, u_{n-1})$ a* finite basis *of x.*

Note. In the sequel, we omit the 'finite' from finite basis whenever there is no risk of confusion. We also write u_j instead of $u_{j \mod n}$ to refer to elements of a finite basis whenever it is convenient to refer to its elements by indexes greater than $n - 1$.

Definition 1.14. *A map $h : \Sigma^* \to \Delta^*$ is a* morphism, *if $h(uv) = h(u)h(v)$ for all $u, v \in \Sigma^*$. A morphism with $|h(a)| = 1$ for all $a \in \Sigma$ is called a* coding.

The notion of morphism extends straightforwardly to infinite words.

Definition 1.15. *An infinite word is called a morphic sequence, if it can be expressed as a coding of a fixed-point of a morphism.*

Definition 1.16. *A 3–sorted algebra* $T = \langle Q, A, B; q_0, \circ, * \rangle$ *is called a* transducer *if* Q, A, B *are finite, nonempty sets, called the* set of states*, the* input alphabet *and the* output alphabet*, respectively,* $q_0 \in Q$ *is called the* initial state*,* $\circ : Q \times A \to Q$ *is a total function called the* transition function *and* $* : Q \times A \to B^*$ *is a total function called the* output function*. We write* $\langle Q, A, B; \circ, * \rangle$ *or even* $\langle Q, A, B; q_0 \rangle$ *if there is no danger of confusion.*

The mappings $\circ$ *and* $*$ *are extended to* $Q \times A^*$ *by defining*

$$q \circ \lambda = q, \qquad q \circ (ua) = (q \circ u) \circ a,$$
$$q * \lambda = \lambda, \qquad q * (ua) = (q * u)((q \circ u) * a),$$

for all $q \in Q$, $(u, a) \in A^* \times A$*. Henceforth, we shall omit parentheses if there is no danger of confusion. So, for example, we will write* $q \circ u * a$ *instead of* $(q \circ u) * a$*.*

We say that T transduces x *to* y *and write* $y = T(x)$ *if* $q_0 * x = y$*.*

A transducer such that $|q * a| = 1$ *for all* $q \in Q$ *and* $a \in A$ *is called a* Mealy machine*.*

1.2 Periodicity of bi-ideals

Since bi-ideals include periodic words, a natural and interesting question to ask is whether a given basis produces a periodic bi-ideal. An answer to this question was given by Buls and Lorencs (2008). In this section we present their results. Section 1.2.1 gives a general condition on a bi-ideal sequence corresponding to a periodic bi-ideal (Theorem 1.23), while Section 1.2.2 gives a combinatorial condition on the basis of such a bi-ideal (Theorem 1.25).

1.2.1 A general condition of periodicity

The following three lemmas are very easy, but they turn out to be extremely useful:

Lemma 1.17. *If* $x = w^\omega$ *and* T *is the minimal period of the word* x*, then* $T \backslash |w|$*, i.e.* T *divides* $|w|$*.*

Proof. Let $n = T|w|$, then both T and $|w|$ are periods of the word $x[0, n)$. Hence (from Theorem 1.3) $t = \gcd(T, |w|)$ is a period of $x[0, n)$. Now we have

$$\forall i \; x[0, n) = x[ni, n(i + 1)).$$

Therefore t is a period of x. Since T is the minimal period of the word x, then $t \geq T \geq \gcd(T, |w|) = t$. Hence $T = \gcd(T, |w|)$, thereby $T \backslash |w|$. □

Lemma 1.18. *If $x = w^\omega = uvy$ and $|w| = |v|$, then $vy = y = v^\omega$.*

Proof. Let $|w| = t$ and $|u| = k + 1$, then $v = x_{k+1}x_{k+2} \dots x_{k+t}$, since $|v| = |w|$. We have $\forall i \; x_{i+t} = x_i$, therefore

$$\forall j \in \overline{1, t} \; \forall s \;\; x_{k+j} = x_{k+j+st} \, .$$

□

Lemma 1.19. *If $\exists u \in A^+ \; ux = x \in A^\omega$, then a word x is periodic with the minimal period $T \backslash |u|$.*

Proof. Let $u = a_1 a_2 \dots a_{t-1}$, where $\forall j \; a_j \in A$, and $y = ux$, then
$\forall i \; x_i = y_{i+t}$. Let

$$y = ux = x \, .$$

Hence

$$\forall i \; y_i = x_i = y_{i+t} \, .$$

This means that y is periodic with a period t. Since $y = x$, then x is periodic with a period t too. Let T is the minimal period of x, then by Lemma 1.17 $T \backslash t$, i.e. $T \backslash |u|$. □

Corollary 1.20. *Let $|v|$ be the minimal period of $x = v^\omega$.*

$$If \qquad v = x[k, k + |v|) \qquad then \qquad |v| \backslash k.$$

Proof. If, for any k, $v = x[k, k + |v|)$, then (see Lemma 1.18)

$$x = x[0, k)v^\omega = x[0, k)x.$$

Hence by Lemma 1.19 $|v| \backslash |x[0, k)| = k$. □

Lemma 1.21. *If there is an integer n such that $v_n u \in v^*$ and $\forall i \in \mathbb{Z}_+ \; (u_{n+i} \in uv^*)$, then*

$$\forall i \in \mathbb{N} \; (v_{n+i} \in v^* v_n) \, .$$

Proof. If $i = 0$ then $v_{n+i} = v_n = \lambda v_n \in v^* v_n$.

Further, we shall prove the lemma by induction on i, i.e., suppose that $v_{n+i} \in v^* v_n$, namely,

$$\exists k \in \mathbb{N} \; (v_{n+i} = v^k v_n) \, .$$

By assumption, $v_n u \in v^*$ and $u_{n+i+1} \in uv^*$, i.e.

$$\exists l \in \mathbb{N}\ (v_n u = v^l)\ \wedge\ \exists m \in \mathbb{N}\ (u_{n+i+1} = uv^m).$$

Hence

$$\begin{aligned} v_{n+i+1} &= v_{n+i}u_{n+i+1}v_{n+i} = (v^k v_n)(uv^m)(v^k v_n) \\ &= v^k(v_n u)v^{m+k}v_n = v^k v^l v^{m+k} v_n \in v^* v_n\,. \end{aligned}$$

We have completed the inductive step. □

Lemma 1.22. *If t is the period of the bi-ideal x and $|v_n| \geq t$, then*

$$\forall i \in \mathbb{Z}_+\ u_{n+1}x = u_{n+i}x\,.$$

Proof. We have $v_{n+i} = v_{n+i-1}u_{n+i}v_{n+i-1}$. Hence, if $i \in \mathbb{Z}_+$ then

$$\forall i \in \mathbb{Z}_+\ \exists v_i'\ v_{n+i} = v_n v_i' v_n\,.$$

Now, by definition of x

$$\begin{aligned} x &= v_n u_{n+1} v_n \ldots \\ x &= v_{n+i}u_{n+i+1}v_{n+i}\ldots = v_n v_i' v_n u_{n+i+1} v_n \ldots \end{aligned}$$

By assumption, x is periodic, therefore

$$x = v^\omega, \quad \text{where} \quad |v| = t\,.$$

Since $v \in \mathrm{Pref}(v_n)$ then by Lemma 1.18

$$\begin{aligned} x &= v_n u_{n+1} x\,, \\ x &= v_n u_{n+i+1} x\,. \end{aligned}$$

Hence $\forall i \in \mathbb{Z}_+\ x = v_n u_{n+i} x$. Thus $\forall i \in \mathbb{Z}_+\ u_{n+1}x = u_{n+i}x$. □

Theorem 1.23. *A bi-ideal x is periodic if and only if*

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

Proof. $\Rightarrow$ Let T be the minimal period of the word x, then $\exists n \in \mathbb{N}\ |v_n| \geq T$. Thus by Lemma 1.22

$$\forall i \in \mathbb{Z}_+ \quad u_{n+1}x = u_{n+i}x\,.$$

Let u be the longest word of the set $\bigcap_{i=1}^{\infty} \mathrm{Pref}(u_{n+i})$ then

$$\forall i \in \mathbb{Z}_+ \exists u'_i \, (u_{n+i} = uu'_i) \, .$$

Particularly, $\exists k \; u_{n+k} = u$. This means that

$$\forall i \in \mathbb{Z}_+ \quad uu'_i x = u_{n+i} x = u_{n+k} x = ux \, .$$

Thus

$$\forall i \in \mathbb{Z}_+ \quad u'_i x = x \, .$$

Hence by Lemma 1.19

$$\forall i \in \mathbb{Z}_+ \quad T \backslash |u'_i| \, .$$

Thereby

$$\forall i \in \mathbb{Z}_+ \quad u'_i \in v^* \, ,$$

where $v = x[0, T)$. Thus

$$\forall i \in \mathbb{Z}_+ \quad u_{n+i} = uu'_i \in uv^* \, .$$

Note

$$x = v_n u_{n+1} v_n \ldots = v_n u u'_1 v_n \ldots$$

Since $u'_1 \in v^*$ and $v \in \mathrm{Pref}(v_n)$, then [Lemma 1.18] $x = v_n ux$. Hence [Lemmma 1.19] $v_n u \in v^*$.

$\Leftarrow$ By Lemma 1.21

$$\forall i \in \mathbb{N} \, \exists k_i \in \mathbb{N} \; v_{n+i} = v^{k_i} v_n \, .$$

Since $\lim\limits_{k \to \infty} |v_k| = \infty$ then $\lim\limits_{i \to \infty} k_i = \infty$. Thus

$$x = \lim_{k \to \infty} v_k = \lim_{i \to \infty} v_{n+i} = \lim_{i \to \infty} v^{k_i} v_n = v^{\omega} \, .$$

□

1.2.2 A combinatorial condition of periodicity

Observation. If all $u_i \in w^*$ for some word $w \neq \lambda$, then the bi-ideal generated by (u_i) is periodic.

The following example demonstrates the converse is not true in general.

Example 1.24. *Let x be the bi-ideal generated by (u_i), where*

$$\begin{aligned} u_0 &= 0, \\ u_1 &= 1, \\ \forall i > 1 \quad u_i &= 00100\,. \end{aligned}$$

Then

$$\begin{aligned} v_0 &= 0, \\ v_1 &= 010, \\ v_2 &= 010\ 00100\ 010, \\ v_3 &= 01000100010\ 00100\ 01000100010, \\ &\ .\quad .\quad . \end{aligned}$$

and $x = \lim\limits_{i\to\infty} v_i = (0100)^\omega$. Thus x is periodic.

Nevertheless, if every u_j appears infinitely often in (u_i), then the converse is valid.

Theorem 1.25. *Let (u_i) be a sequence of words, which contains every u_j infinetely often. The bi-ideal x generated by (u_i) is periodic if and only if*

$$\exists w \forall i\ u_i \in w^*\,.$$

Proof. $\Rightarrow$ Let x be a periodic bi-ideal, then by Theorem 1.23

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

Hence by Lemma 1.21 $|v|$ is the period of x. Therefore we can assume that $|v|$ is the minimal period of x and $|u| < |v|$. Since the sequence (u_i) contains every u_j infinitely often then by Theorem 1.23 $\forall i \in \mathbb{N}\ (u_i \in uv^*)$.

Now suppose that $u_i = u$ for all $i < m$ but $u_m = uv^k$, where $k > 0$. Then there exist $\alpha \in \mathbb{Z}_+$ and y such that

$$x = u^\alpha v^k y\,.$$

(i) If $u = \lambda$ then $\forall i\ u_i \in v^*$.

(ii) Otherwise $u \neq \lambda$. Then (Corollary 1.20) $|v| \backslash \alpha|u|$. Hence, there exists $\beta \in \mathbb{Z}_+$ such that $\alpha|u| = \beta|v|$. Thus $x = v^\omega = u^\omega$. Contradiction, since $|u| < |v|$ and $|v|$ is the minimal period of x.

$\Leftarrow$ See Observation. $\square$

Now we turn our attention to the problem of effectiveness.

Theorem 1.26. *A bi-ideal x generated by $(u_0, u_1, \ldots, u_{m-1})$ is periodic if and only if*

$$\exists w \forall i \in \overline{0, m-1} \;\; u_i \in w^* \,.$$

Proof. As a corollary from Definition 1.12 and Theorem 1.25. □

This theorem gives a method to generate aperiodic bi-ideals. Let

$$(u_0, u_1, \ldots, u_{m-1})$$

be any m-tuple chosen at random. Let v be any shortest word from the set

$$\{u_0, u_1, \ldots, u_{m-1}\}$$

and w be the shortest prefix of v such that $v \in w^+$. If there exists u_i such that $u_i \notin w^*$ then the bi-ideal generated by $(u_0, u_1, \ldots, u_{m-1})$ is not periodic. This can be easily checked by a deterministic algorithm.

2 Cryptographic potential of finitely generated bi-ideals

In this chapter possible uses of finitely generated bi-ideals are explored. Section 2.1 is based on work done for my masters thesis and shows a method of selecting bases for bi-ideals in a manner such that the resulting bi-ideal has asymptotically uniformly distributed factors up to a given length. In section 2.2 a construction for an aperiodic pseudo random number generator with good statistical properties, based on finitely generated bi-ideals is given.

2.1 Letter frequencies in finitely-generated bi-ideals

2.1.1 The frequency test

Let's look at the prefix of a bit sequence $\{x_n\}$

$$x = (x_1, x_2, \ldots, x_{N+\nu-1})$$

This prefix has $N = |x| - \nu + 1$ overlapping sub-sequences of length ν.

Consider a specific bit sequence of the length ν.

$$s = (s_1, s_2, \ldots, s_\nu)$$

We can denote the event of the m-th sub-sequence of $\{x_n\}$ being equal to s with

$$D_s^m(x) = \{(x_m, x_{m+1}, \ldots, x_{m+\nu-1}) = s\}$$

If $\{x_n\}$ is indistinguishable from an i.i.d. bit-sequence, then

$$E(I(D_s^m(x))) = 2^{-\nu},$$

where I is the indicator function and E denotes the expected value. Or, if we denote the number of occurrences of the sequence s in x with $|x|_s$,

$$E(|x|_s) = 2^{-\nu}N.$$

For a broader coverage see Neuenschwander (2004).

2.1.2 Finitely generated bi-ideals and the frequency test

A convergence theorem

Let u and w be finite words. Then we can denote

(i) $|w|_u = |\{(u', u, u'')|u'uu'' = w\}|$.

$|w|_u$ is the count of different ways u is a factor of w.

(ii) We will call the number $\alpha(w, u) = \dfrac{|w|_u}{|w| - |u| + 1}$ the *relative frequency* of u in w. By this definition $0 \leq \alpha(w, u) \leq 1$.

Suppose, x is a bi-ideal generated by the sequence (u_i), then we can denote

(iii) $\alpha_n(u) = \alpha(v_n, u)$, where v_n is the n-th element of the bi-ideal sequence from Definition 1.8, where $x = \lim_{i\to\infty} v_i$.

Lemma 2.1. *If x is a restricted bi-ideal, then*

$$\forall l \in \mathbb{N}\, \forall \varepsilon > 0\, \exists \delta \in \mathbb{N}\, \forall u \in A^*[|u| = l \Rightarrow \forall n \geq \delta\ |\alpha_n(u) - \alpha_\delta(u)| \leq \varepsilon].$$

Proof. Suppose, the sequence (u_i), generates an l_x–restricted bi-ideal x. Let's consider the bi-ideal sequence (v_i) generated by (u_i) as per Definition 1.7. Then from Definition 1.8 each v_i is a prefix of the bi-ideal x, and $|v_j| > |v_i|$, when $j > i$.

Let's denote:

$$\begin{aligned} l_i &= |v_{\delta+i}|_u, \quad i \geq 0 & \qquad l'_i &= l_i - 2l_{i-1}, \quad i \geq 1; \\ m_i &= |v_{\delta+i}| - l + 1, \quad i \geq 0 & \qquad m'_i &= m_i - 2m_{i-1}, \quad i \geq 1; \end{aligned}$$

$$\alpha_i = \alpha(v_{\delta+i}, u) = \frac{|v_{\delta+i}|_u}{|v_{\delta+i}| - |u| + 1} = \frac{|v_{\delta+i}|_u}{|v_{\delta+i}| - l + 1} = \frac{l_i}{m_i}, \quad i \geq 0,$$

where $l = |u|$.

We will choose δ such, that $\dfrac{l_x + l - 1}{m_0} < \varepsilon$. $(m_0 = |v_\delta| - l + 1)$

Let's asses α_i, $i \geq 1$:

$$\begin{aligned} \alpha_i &= \frac{l_i}{m_i} = \frac{2l_{i-1} + l'_i}{2m_{i-1} + m'_i} = \frac{2l_{i-1}}{2m_{i-1} + m'_i} + \frac{l'_i}{2m_{i-1} + m'_i} \\ &= \alpha_{i-1}\frac{2}{2 + \frac{m'_i}{m_{i-1}}} + \frac{l'_i}{2m_{i-1} + m'_i} = \alpha_{i-1}\frac{1}{1 + \frac{m'_i}{2m_{i-1}}} + \frac{l'_i}{2m_{i-1} + m'_i}. \end{aligned}$$

From Definition 1.7, $v_{i+1} = v_i u_{i+1} v_i, \quad i \geq 1$, therefore

$$|v_{i+1}| = 2|v_i| + |u_{i+1}| \geq 2|v_i|, \tag{2.1}$$

and $|v_{\delta+i}| \geq 2^i |v_\delta|$. From here

$$m_{i-1} = |v_{\delta+i-1}| - l + 1 \geq 2^{i-1}|v_\delta| - l + 1 \geq 2^{i-1}(|v_\delta| - l + 1) = 2^{i-1} m_0. \tag{2.2}$$

Now consider,

$$\begin{aligned} m'_i &= m_i - 2m_{i-1} = |v_{\delta+i}| - l + 1 - 2(|v_{\delta+i-1}| - l + 1) \\ &= |v_{\delta+i}| - l + 1 - 2|v_{\delta+i-1}| + 2l - 2, \end{aligned}$$

from (2.1), $|v_{\delta+i}| - |v_{\delta+i-1}| = |u_{\delta+i}|$, therefore

$$|v_{\delta+i}| - l + 1 - 2|v_{\delta+i-1}| + 2l - 2 = |u_{\delta+i}| + l - 1.$$

But because the bi-ideal is l_x restricted, $|u_{\delta+1}| \leq l_x$ and $m'_i \leq l_x + l - 1$ therefore, also considering (2.2):

$$\frac{m'_i}{2m_{i-1}} \leq \frac{l_x + l - 1}{2^i m_0} \leq \frac{\varepsilon}{2^i}.$$

Now we can remember, that $1 \geq 1 - a^2 = (1 - a)(1 + a)$, and, if $1 + a > 0$,

$$\frac{1}{1+a} \geq 1 - a,$$

so that we can write,

$$\begin{aligned} \alpha_i &= \alpha_{i-1} \frac{1}{1 + \frac{m'_i}{2m_{i-1}}} + \frac{l'_i}{2m_{i-1} + m'_i} \geq \alpha_{i-1} \left(1 - \frac{m'_i}{2m_{i-1}}\right) \\ &\geq \alpha_{i-1} \left(1 - \frac{\varepsilon}{2^i}\right) = \alpha_{i-1} - \frac{\alpha_{i-1}\varepsilon}{2^i} \geq \alpha_{i-1} - \frac{\varepsilon}{2^i}. \end{aligned}$$

Now let's look at l'_i:

$$l'_i = l_i - 2l_{i-1} = |v_{\delta+i}|_u - 2|v_{\delta+i-1}|_u$$

But by Definition 1.7, $v_{\delta+i} = v_{\delta+i-1} u_{\delta+i} v_{\delta+i-1}$, and in $v_{\delta+i-1}$ there are l_{i-1} factors equal to u. We know, that of the $2m_{i-1}$ factors with the length l corresponding to the $v_{\delta+i-1}$ precisely $2l_{i-1}$ are equal to u. This means, that there can be at most $l_i \leq 2l_{i-1} + m_i - 2m_{i-1}$ factors equal to u. And thus,

$$l'_i \leq m_i - 2m_{i-1} \leq m'_i \leq l_x + l - 1$$

Also, from (2.2):

$$2m_{i-1} + m'_i \geq 2m_{i-1} \geq 2^i m_0.$$

From this

$$\begin{aligned} \alpha_i &= \alpha_{i-1} \frac{1}{1 + \frac{m'_i}{2m_{i-1}}} + \frac{l'_i}{2m_{i-1} + m'_i} \\ &\leq \alpha_{i-1} + \frac{l_x + l - 1}{2^i m_0} \leq \alpha_{i-1} + \frac{\varepsilon}{2^i}. \end{aligned}$$

We have assessed

$$\alpha_{i-1} - \frac{\varepsilon}{2^i} \leq \alpha_i \leq \alpha_{i-1} + \frac{\varepsilon}{2^i},$$

so that,

$$\alpha_0 - \varepsilon \sum_{j=1}^{i} 2^{-j} \leq \alpha_i \leq \alpha_0 + \varepsilon \sum_{j=1}^{i} 2^{-j}.$$

And, because $\sum_{j=1}^{\infty} 2^{-j} = 1$, we can write

$$\alpha_0 - \varepsilon \leq \alpha_i \leq \alpha_0 + \varepsilon.$$

If we remember, that $\alpha_0 = \alpha_\delta(u)$ and $\alpha_i = \alpha_{\delta+i}(u)$, we can conclude, that the lemma is proved. □

Theorem 2.2. *If x is a restricted bi-ideal, and V_k denotes a prefix of x with length k, then*

$$\forall l \in \mathbb{N}\, \forall \varepsilon > 0\, \exists K \in \mathbb{N}\, \forall u \in A^*\, [|u| = l \Rightarrow \forall k \geq K\ |\alpha(V_K, u) - \alpha(V_k, u)| \leq \varepsilon].$$

Proof. We will use denotations similar to those, used in the proof of the preceding lemma:

$$\begin{aligned} l_i &= |v_i|_u, \quad i \geq 0 \\ m_i &= |v_i| - l + 1, \quad i \geq 0 \\ \alpha_i &= \alpha(v_i, u) = \frac{l_i}{m_i}, \quad i \geq 0, \end{aligned}$$

where $l = |u|$, and (v_i) is the bi-ideal sequence associated with x as by Definition 1.8. Also, we assume, that the bi-ideal is l_x restricted.

We select a parameter n, such that:

$$\forall i \geq 1 \quad |\alpha_{n+i}(u) - \alpha_n(u)| < \frac{\varepsilon}{4} \tag{2.3}$$

$$\frac{l_x + l}{m_n} < \frac{\varepsilon}{4} \tag{2.4}$$

Note, that the first condition can be satisfied according to Lemma 2.1.

Then we select a parameter $g > n$, such, that

$$\frac{m_n + l_x + l}{m_g} < \frac{\varepsilon}{4} \tag{2.5}$$

According to Lemma 2.1 and the way n and g were chosen

$$\alpha_n(u) - \frac{\varepsilon}{4} < \alpha_g(u) < \alpha_n(u) + \frac{\varepsilon}{4}.$$

We introduce a function $j : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$, and denote,

$$v_{g,h,n} = v_g u_{j(g,1)} v_n u_{j(g,2)} v_n \dots u_{j(g,h)} v_n,$$

such, that $v_{g,h,n} \in Pref\ x$.

Now we can introduce corresponding denotations:

$$l_{g,h,n} = |v_{g,h,n}|_u$$
$$m_{g,h,n} = |v_{g,h,n}| - l + 1$$
$$\alpha_{g,h,n} = \frac{l_{g,h,n}}{m_{g,h,n}} = \alpha(v_{g,h,n}, u)$$

Considering the construction of $v_{g,h,n}$ and that the bi-ideal is l_x restricted,

$$m_g + hm_n \le m_{g,h,n} \le m_g + hm_n + hl_x + hl$$

Let's asses $l_{g,h,n}$. It is obvious from the construction, that

$$l_{g,h,n} \ge l_g + hl_n$$

To asses the upper bound of $l_{g,h,n}$, we have to remember, that the word $v_{g,h,n}$ has a total number of $m_{g,h,n}$ factors with length l, however the words v_g and v_n have correspondingly m_g and m_n factors with length l. That means, that the word $v_{g,h,n}$ has a maximum of

$$m_g + hm_n + hl_x + hl - m_g - hm_n = h(l_x + l)$$

factors of length l, that we don't know being equal to u or not. Therefore,

$$l_{g,h,n} \le l_g + h(l_n + l_x + l).$$

Now we can asses $\alpha_{g,h,n}$. We will start with the lower limit:

$$\alpha_{g,h,n} = \frac{l_{g,h,n}}{m_{g,h,n}} \ge \frac{l_g + hl_n}{m_g + hm_n + hl_x + hl} = \frac{\alpha_g m_g + h\alpha_n m_n}{m_g + hm_n + hl_x + hl}$$

according to condition (2.3)

$$\begin{aligned}\frac{\alpha_g m_g + h\alpha_n m_n}{m_g + hm_n + hl_x + hl} &\geq \frac{(\alpha_n - \frac{\varepsilon}{4})m_g + h\alpha_n m_n}{m_g + hm_n + hl_x + hl} \\ &= \alpha_n \frac{1}{1 + \frac{hl_x + hl}{m_g + hm_n}} - \frac{\varepsilon}{4}\frac{m_g}{m_g + hm_n + hl_x + hl}\end{aligned}$$

according to condition (2.4)

$$\frac{hl_x + hl}{m_g + hm_n} \leq \frac{\varepsilon}{4},$$

therefore

$$\alpha_n \frac{1}{1 + \frac{hl_x + hl}{m_g + hm_n}} - \frac{\varepsilon}{4}\frac{m_g}{m_g + hm_n + hl_x + hl} \geq \alpha_n \frac{1}{1 + \frac{\varepsilon}{4}} - \frac{\varepsilon}{4} \geq \alpha_n - \frac{\varepsilon}{2},$$

and

$$\alpha_{g,h,n} \geq \alpha_n - \frac{\varepsilon}{2} \tag{2.6}$$

We now have to asses the upper limit of $\alpha_{g,h,n}$:

$$\alpha_{g,h,n} = \frac{l_{g,h,n}}{m_{g,h,n}} \leq \frac{l_g + hl_n + h(l_x + l)}{m_g + hm_n} = \frac{\alpha_g m_g + h\alpha_n m_n + h(l_x + l)}{m_g + hm_n}$$

we can again use the condition (2.3)

$$\begin{aligned}\frac{\alpha_g m_g + h\alpha_n m_n + h(l_x + l)}{m_g + hm_n} \leq \frac{(\alpha_n + \frac{\varepsilon}{4})m_g + h\alpha_n m_n + h(l_x + l)}{m_g + hm_n} = \\ = \alpha_n + \frac{\varepsilon}{4}\frac{m_g}{m_g + hm_n} + \frac{h(l_x + l)}{m_g + hm_n}\end{aligned}$$

according to the condition (2.4)

$$\frac{h(l_x + l)}{m_g + hm_n} \leq \frac{\varepsilon}{4},$$

and, therefore

$$\alpha_{g,h,n} \leq \alpha_n + \frac{\varepsilon}{2}. \tag{2.7}$$

Finally, let's look at V_i — a prefix of x with a length of $i > |v_g|$. We can find such a h, that $V_i = v_{g,h,n}w$, where $w \in Pref(u_{j(g,h+1)}v_n)$.

Let's asses $\alpha(V_i, u) = \frac{|V_i|_u}{|V_i| - l + 1}$.

It is clear from the way we selected i (and implicitly h), that

$$m_{g,h,n} \leq |V_i| - l + 1 \leq m_{g,h,n} + m_n + l_x + l,$$

also, it is obvious, that

$$l_{g,h,n} \leq |V_i|_u \leq l_{g,h,n} + m_n + l_x + l.$$

We can asses the lower limit for $\alpha(V_i, u)$:

$$\alpha(V_i, u) \geq \frac{l_{g,h,n}}{m_{g,h,n} + m_n + l_x + l} = \alpha_{g,h,n} \frac{1}{1 + \frac{m_n + l_x + l}{m_{g,h,n}}}$$

according to the condition (2.5), and considering, that $m_{g,h,n} \geq m_g$,

$$\frac{m_n + l_x + l}{m_{g,h,n}} \leq \frac{\varepsilon}{4}$$

therefore,

$$\alpha_{g,h,n} \frac{1}{1 + \frac{m_n + l_x + l}{m_{g,h,n}}} \geq \alpha_{g,h,n} \frac{1}{1 + \frac{\varepsilon}{4}} \geq \alpha_{g,h,n} - \frac{\varepsilon}{4}$$

using (2.6) and (2.3)

$$\alpha(V_i, u) \geq \alpha_{g,h,n} - \frac{\varepsilon}{4} \geq \alpha_n - \frac{3\varepsilon}{4} \geq \alpha_g - \varepsilon$$

The last thing we have to do, is to asses the upper limit of $\alpha(V_i, u)$:

$$\alpha(V_i, u) \leq \frac{l_{g,h,n} + m_n + l_x + l}{m_{g,h,n}} = \alpha_{g,h,n} + \frac{m_n + l_x + l}{m_{g,h,n}}$$

according to the condition (2.5)

$$\frac{m_n + l_x + l}{m_{g,h,n}} \leq \frac{\varepsilon}{4},$$

therefore,

$$\alpha_{g,h,n} + \frac{m_n + l_x + l}{m_{g,h,n}} \leq \alpha_{g,h,n} + \frac{\varepsilon}{4},$$

and, using (2.7) and (2.3)

$$\alpha(V_i, u) \leq \alpha_{g,h,n} + \frac{\varepsilon}{4} \leq \alpha_n + \frac{3\varepsilon}{4} \leq \alpha_g + \varepsilon.$$

So we can write:

$$\alpha_g - \varepsilon \ \leq \alpha(V_i, u) \leq \alpha_g + \varepsilon$$

If we examine the conditions of the theorem, we see, that it is proved, and that $K = |v_g|$. □

'Good' bases for the frequency test

For a bit-sequence to be indistinguishable from i.i.d bit-sequences, each of the test words of a given length ν, have to appear an equal number of times. Some deviations are of course permitted, depending on the statistical test we use, to check this property.

The suggested method is as follows:

1. Choose any word of length $\nu - 1$. This word, denoted by a, will be a prefix for the generated base words.

2. The base words are found in the form ab such, that all of the test words of the length ν would appear an equal number of times in the word aba.

We will call the base words yielded by this method *good base words for a test length of* ν.

For example:

$$\underbrace{101}_{a}\underbrace{0010110000111}_{b}\underbrace{101}_{a}$$

We can see, that each of the test words of the length 4 $(0000, 0001, 0010, \ldots, 1111)$ appear in the word aba exactly once. In this case, $ab = 1010010110000111$ is a good base word for the test length of 4. Although we do not currently have a precise estimate of the number of good test words, a full search reveals, that there are 32 good base words with this prefix for the test length of 4, that contain each of the test words exactly once, and 209952, that contain each of the test words exactly twice.

Lemma 2.3. *Given a long enough sequence of bits is tested, a restricted bi-ideal, generated from good base words for a test length of* ν*, will be indistinguishable from an i.i.d. bit-sequence using the frequency test with a test word length of* ν*.*

Proof. Let's consider the bi-ideal x. It can be written as:

$$x = u_0 u_1 u_0 u_2 u_0 u_1 \ldots$$

If we introduce a function $j : \mathbb{N} \to \mathbb{N}$, such that $j(i)$ is the index of the i-th base word in the bi-ideal x. Then, $j(0) = 0; j(1) = 1; j(2) = 0; j(3) = 2$, and so on.

Let's denote:

$$x_i = u_0 u_1 u_0 \ldots u_{j(i)} \# Pref_{\nu-1}(u_{j(i+1)}),$$

where $Pref_n(u)$ denotes the prefix of the word u, of length n. Using the notation of the Definition, if the word u would be expressed as ab, where a is a prefix of the length $\nu - 1$, it is clear, that $\forall i,\ Pref_{\nu-1}(u_i) = a$.

Let's consider $x_0; x_1$, and so on.

According to our definition of the good base words, and considering

$$\forall u_i, u_j\ Pref_{\nu-1} u_i = Pref_{\nu-1} u_j,$$

each test-word with a length of ν, will appear in x_0 an equal number of times.

Let's compare x_1 and x_0. It is obvious, that each test word with the length ν will appear in x_1 the same number of times, as it appears in x_0, plus so many times, as it appears in the word $u_1\#Pref_{\nu-1}u_2$. But considering $\forall u_i, u_j \; Pref_{\nu-1}u_i = Pref_{\nu-1}u_j$, and that u_1, is a good base word as well, it becomes obvious, that each of the test-words appears in x_1, an equal number of times as well. It is clear, that this can be shown for any x_k in a similar fashion.

It is clear, that any deviation from this occurs only, when the prefix does not equal to one of the values of x_i. However, the maximum deviation for any given test-word, will never exceed the maximum number of times the test-word appears in the longest base word of the bi-ideal. And thus, if we look at the relative frequency of the test-word, we see, that it is inversely proportional to the length of the bi-ideal. This means, that for a sufficiently long prefix of x, the bi-ideal will not be distinguishable from an i.i.d. bit-sequence using the frequency test, with a test-word length of ν. □

Lemma 2.4. *A good base word for a test-length of ν, is a good base word for all test-lengths, smaller then ν.*

Proof. Suppose, we have a good base word for the test-length of ν. Obviously, $|u| = 2^\nu k$.

From the definition of the good base words, it is clear, that each of the test words, with a length of ν, appears in the word $u\#Pref_{\nu-1}(u)$, exactly k times.

Let's consider test-words with a length of $\nu - 1$. If or assumption is correct, each of these test-words have to appear in $u\#Pref_{\nu-2}(u)$ exactly $2k$ times. Let's assume the opposite. Then there must be at least one test word v, with a length of $\nu - 1$, that will appear in $u\#Pref_{\nu-2}(u)$ at least $2k + 1$ times.

Let's examine each of the occurrences of v in $u\#Pref_{\nu-2}(u)$. Considering, that $u\#Pref_{\nu-1}(u)$ is one bit longer then $u\#Pref_{\nu-2}(u)$, we can look at each v plus the next bit. It is clear, that this way we will have constructed a word of the length ν, for each occurrence of v, that will be a sub-word of $u\#Pref_{\nu-1}(u)$.

But, considering, that we can only have a 0 or 1 following v, it is clear that either $v\#0$ or $v\#1$ will appear in the word $u\#Pref_{\nu-1}(u)$ at least $k + 1$ times. But this would mean, that u is not a good base word for the length ν. Thus we have a contradiction.

It is clear, that similarly we can show the same for the lengths $\nu - 2$, $\nu - 3$, and so on. □

Theorem 2.5. *A restricted bi-ideal generated from good base words for the length ν will be indistinguishable from an i.i.d. bit-sequence, using test words with a length of up to ν, given a long enough bit-sequence.*

Proof. The proof of the theorem obviously follows from the Lemmas 2.3 and 2.4. □

2.2 Generation of aperiodic pseudo-random sequences with good statistical properties

2.2.1 Introduction

This section presents work, that is a joint effort by the members of the seminar "combinatorics on words" at the University of Latvia, of which I am a member. This work has been published in Berzina et al. (2011).

As of today, the most convenient and reliable way of generating random symbols for stochastic simulations appears to be via deterministic algorithms with a solid mathematical basis. These algorithms produce sequences of symbols which are, in fact, not random at all, but seem to behave as if the symbools were chosen independently at random.

We are interested in methods that generate aperiodic sequences. One method for obtaining aperiodic sequences is to use the simplest chaotic system — the logistic map. In 1982 Oishi and Inoue 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. Recently, Hu et. al. (2009) proposed a true random number generator by combining congruential methods with prime numbers and higher order composition of logistic maps. For more information of using chaotic systems in generation of pseudo-random sequences see e.g. Patidar and Sud (2009), Phatak and Rao (1995).

Here we propose a method to generate aperiodic pseudo-random symbol sequences based on modification of the shrinking generator, which was introduced by Coppersmith et al. (1994) and is still considered a secure pseudo-random symbol generator. Normally, a shrinking generator uses two pseudo-random bit-sequences produced by LFSR's (see, e.g., Schneier (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).

A pseudo-random symbol generator can be created by substituting the S-sequence by an aperiodic sequence — a finitely generated bi-ideal. We conjecture, that for most non-trivial cases the resulting pseudo-random sequence is aperiodic. The resulting pseudo-random sequence has good statistical properties as indicated by the Diehard test suite (see Section 2.2.2).

We show two approaches for generating aperiodic pseudo-random number sequences using our modified shrinking generator. First, 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 aperiodic. Second, we show that there are what we call universal bi-ideals — finitely generated bi-ideals that generate aperiodic 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.

Section 2.2.2 is devoted to the selection of finitely generated bi-ideals given an A-sequence, while in Section 2.2.3 we show that there are infinitely many universal bi-ideals.

2.2.2 Aperiodic 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 aperiodic. Afterwards, we shortly analyse test results.

Definition 2.6. *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.

Construction

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

Theorem 2.7. *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 bi-ideal with basis $\langle u_0, u_1, ..., u_{m-1}\rangle$ is aperiodic.*

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

Example 2.8. *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 2.9. *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 condition 2.7 is satisified, e.g., the bi-ideals used as the S-sequences are aperiodic, 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 an aperiodic 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 2.10. *If $x \in \{0,1\}^\omega$ is a bi-ideal generated by $\langle u_0, u_1, ..., u_{m-1}\rangle$, then $\forall p, T \in \mathbb{N}\ \overset{\infty}{\exists}\alpha, \beta \in \mathbb{N}, \alpha \neq \beta$:*

$$|v_{\alpha m-1}| \equiv |v_{\beta m-1}| \ (mod\ p), \tag{2.8}$$

$$|v_{\alpha m-1}|_1 \equiv |v_{\beta m-1}|_1 \ (mod\ T), \tag{2.9}$$

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:

$$\forall k \geq 1 \quad A_t = \left\{ v_{km-1} \middle|\ |v_{km-1}| \equiv t \pmod{p} \right\}. \tag{2.10}$$

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 (2.8) holds.

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

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

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 (2.8) and (2.9) hold.

□

Lemma 2.11. *(see, e.g., Buls and Lorencs (2008)) 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 2.12. *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 a 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 on contrary that the shrunk sequence is ultimately periodic, e.g., $z = v'v^\omega$ (where $|v'| = T_1$ and $|v| = T$). Then by lemma 2.10 we can choose $\alpha, \beta \in \mathbb{N}$ $(\alpha < \beta)$ such that

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

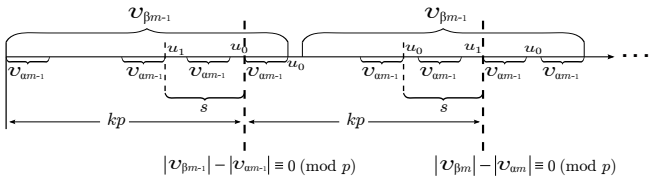

Figure 2.1: Structure of the bi-ideal y.

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

$$|v_{\beta m-1}| - |v_{\alpha m-1}| = kp \tag{2.11}$$

$$|v_{\beta m-1}|_1 - |v_{\alpha m-1}|_1 = k_1 T \tag{2.12}$$

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 (2.11) and (2.12) and using lemma 2.11 we obtain (see Figure 2.1)

$$\begin{aligned} |y[|v_{\beta m-1}| - |v_{\alpha m-1}| + 1, |v_{\beta m}| - |v_{\alpha m}|]| = \\ = |v_{\beta m}| - |v_{\alpha m}| - |v_{\beta m-1}| + |v_{\alpha m-1}| = \\ = 2|v_{\beta m-1}| + |u_0| - 2|v_{\alpha m-1}| - \\ -|u_0| - |v_{\beta m-1}| + |v_{\alpha m-1}| = \\ = |v_{\beta m-1}| - |v_{\alpha m-1}| = kp \end{aligned} \tag{2.13}$$

and

$$\begin{aligned} |y[|v_{\beta m-1}| - |v_{\alpha m-1}| + 1, |v_{\beta m}| - |v_{\alpha m}|]|_1 = \\ = |v_{\beta m-1}|_1 - |v_{\alpha m-1}|_1 = k_1 T. \end{aligned} \tag{2.14}$$

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 (2.13) and (2.14) we obtain

$$S_{y'}(x') = v' z[T_1 + 1, k_1 T], \tag{2.15}$$

$$S_{y''}(x'') = v'' z[k_1 T + T_1 + 1, 2k_1 T], \tag{2.16}$$

where $|v'| = |v''| = T_1$. Next, by (2.15), (2.16), $|v_{\alpha m-1}|_1 > T_1$ and from the assumption that z is ultimately periodic it follows that

$$z[T_1 + 1, k_1 T] = z[k_1 T + T_1 + 1, 2k_1 T], \tag{2.17}$$

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

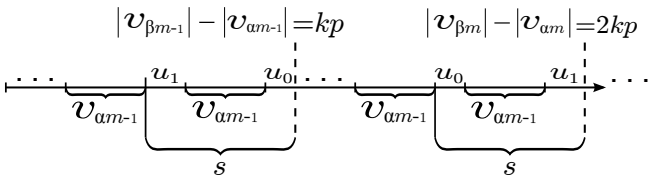

Figure 2.2: Structure of the bi-ideal y.

that

$$S_{y[kp-s+1,kp]}\big(x[kp-s+1,kp]\big) = \\ = S_{y[2kp-s+1,2kp]}\big(x[2kp-s+1,2kp]\big), \quad (2.18)$$

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 (2.18) does not hold, hence proving the existence of a finitely generated bi-ideal y, such that the shrunk word $z = S_y(x)$ is not ultimately periodic.

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

$$\exists i \in \overline{2,p} : (u[i-1] = a \wedge \forall j \in \overline{i,p}\ u[j] = \overline{a})), \quad (2.19)$$

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

$$u_0 = u'10w,\ u_1 = u''01w, \quad (2.20)$$

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 (2.19) and (2.20) $z[k_1T - \gamma] = a$ but $z[2k_1T - \gamma] = \overline{a}$. Thus (2.17) and (2.18) do not hold. Hence z is not ultimately periodic. Since $u', u'' \in \{0,1\}^*$ are arbitrary finite words over alphabet $\{0,1\}$, there exist 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 2.13. *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. Since each periodic word is also ultimately periodic, the proof follows directly from the proof of the Proposition 2.12. □

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 passes *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 lengths around 2KB (the base words were initialized by another 32 bit LFSR) as the S-sequence.

2.2.3 Universal Bi-ideals

In Section 2.2.2 we showed how it is possible to construct apoeriodic S-sequences for each periodic A-sequence such that the resulting shrunk words are aperiodic. 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 aperiodic bi-ideals (as S-sequences) such that for each non-trivial A-sequence the resulting shrunk word would be aperiodic. In Proposition 2.17 we prove the existence of such bi-ideals.

Definition 2.14. *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 2.15. *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 2.16. *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. We denote by (v_n) 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 $j \equiv k+1 \pmod m$ and both $00 \notin \mathrm{F}(v_k)$ and $00 \notin \mathrm{F}(u_j)$, and $1 = v_0 \in \mathrm{Pref}(v_k) \cap \mathrm{Suff}(v_k)$ (by lemma 2.11), then $00 \notin \mathrm{F}(v_{k+1})$.

□

Proposition 2.17. *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 $\langle u_0, u_1, ..., u_{m-1}\rangle$. Assume on 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 2.10, 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

$$|v_{\beta m-1}| - |v_{\alpha m-1}| = kp, \tag{2.21}$$

$$|v_{\beta m-1}|_1 - |v_{\alpha m-1}|_1 = k_1 T. \tag{2.22}$$

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

Now, similarly to the proof of Proposition 2.12, we observe that $v_{\alpha m} = v_{\alpha m-1} 1 v_{\alpha m-1}$ and $v_{\beta m} = v_{\beta m-1} 1 v_{\beta m-1}$ and, therefore, from (2.21), (2.22) and using lemma 2.11 we obtain (see Figure 2.3)

$$\begin{aligned} |y[|v_{\beta m-1}| - |v_{\alpha m-1}| + 1, |v_{\beta m}| - |v_{\alpha m}|]| = \\ = |v_{\beta m-1}| - |v_{\alpha m-1}| = kp \end{aligned} \tag{2.23}$$

and

$$\begin{aligned} |y[|v_{\beta m-1}| - |v_{\alpha m-1}| + 1, |v_{\beta m}| - |v_{\alpha m}|]|_1 = \\ = |v_{\beta m-1}|_1 - |v_{\alpha m-1}|_1 = k_1 T. \end{aligned} \tag{2.24}$$

Now, from the periodicity of x and the equations (2.23) and (2.24) we obtain

$$\begin{aligned} x[kp - |v_{\alpha m-1}|, kp - 1] = \\ = x[2kp - |v_{\alpha m-1}|, 2kp - 1], \end{aligned} \tag{2.25}$$

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

and,

$$|y[kp - |v_{\alpha m-1}|, kp]|_1 = |y[2kp - |v_{\alpha m-1}| - 1, 2kp]|_1. \tag{2.27}$$

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

$$x' = x[kp - \ell, kp - 1],$$

$$x'' = x[2kp - \ell - 1, 2kp - 2] = x[kp - \ell - 1, kp - 2],$$

$$y' = v_{\alpha m-1},$$

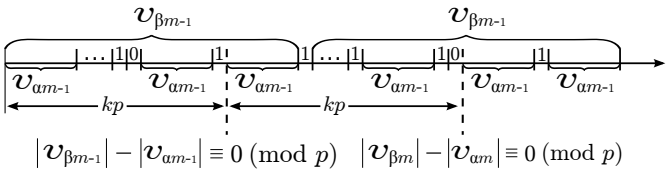

Figure 2.3: Structure of the bi-ideal y.

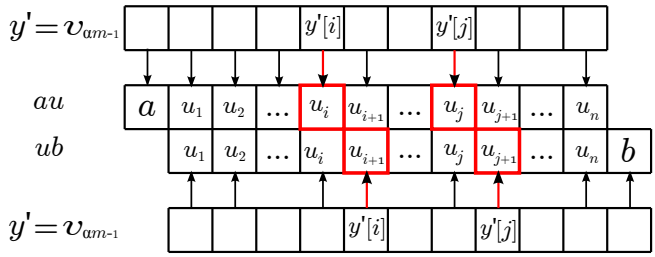

Figure 2.4: Structure of the bi-ideal y.

then from here, (2.25), (2.26) and (2.27) and using our assumption that z is ultimately periodic we obtain

$$S_{v_{\alpha m-1}}(x[kp-l,kp-1]) = \\ = S_{v_{\alpha m-1}}(x[kp-l-1,kp-2]). \quad (2.28)$$

If we further set $x[kp-l-1,kp-1] = avb = v' = v'_1v'_2\dots v'_{l+1}$, then

$$S_{v_{\alpha m-1}}(av) = S_{v_{\alpha m-1}}(vb), \quad (2.29)$$

but $av \neq vb$ from lemma 2.15. From here

$$\exists i > 1 \forall j \leq i: \ v'[j-1] = v'[j] \wedge v'[i] \neq v'[i+1], \quad (2.30)$$

but from (2.29) it follows that

$$\forall s \in \overline{1,l}\ S_{y'[1,s]}(v'[1,s]) = S_{y'[1,s]}(v'[2,s+1]). \quad (2.31)$$

Observe that if i is the index mentioned in (2.30) and $y'[i] = 1$ then from (2.31) equation (2.29) does not hold (see Figure 2.4). Thus $y'[i] = 0$. Moreover, since (2.31) holds for all $s \in \{1,2,\dots,l\}$ then

$$\forall t \in \overline{1,l-1}: \ v'[t] \neq v'[t+1] \Rightarrow y'[t] = 0. \quad (2.32)$$

Since $|alph(u)| = 2$ and $\ell > p$ there exists an index $t_0 < p$ such that (2.32) holds. From this, (2.32) and the periodicity of x we get that for all $t \in \{1,2,\dots,p-1\}$ and for all $\mu \in \mathbb{N}$

$$(v'[t] \neq v'[t+1] \wedge t + \mu p < l) \Rightarrow y'[t+\mu p] = y'[t] = 0, \quad (2.33)$$

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

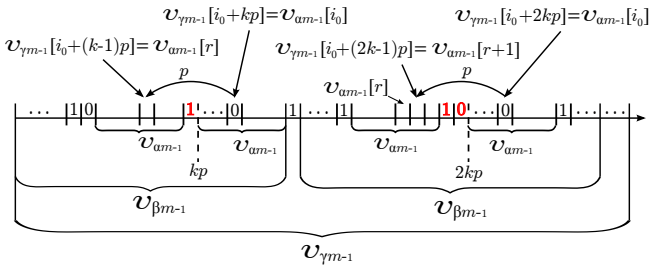

Figure 2.5: Structure of the bi-ideal with basis $\langle 1, 10\rangle$.

considering $|alph(u)| = 2$ and $|v_{\gamma m-1}| > \ell > p$, there exists an index $i_0 < p$ such that for all $\eta \in \mathbb{N}$

$$i_0 + \eta p \leq |v_{\gamma m-1}| \Rightarrow v_{\gamma m-1}[i_0] = v_{\gamma m-1}[i_0 + \eta p] = 0. \tag{2.34}$$

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

$$i_0 + \eta p \leq |v_{\alpha m-1}| \Rightarrow v_{\alpha m-1}[i_0] = v_{\alpha m-1}[i_0 + \eta p] = 0. \tag{2.35}$$

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 (2.21), (2.22) and (2.34) we obtain

$$v_{\gamma m-1}[i_0] = v_{\gamma m-1}[i_0 + kp] = v_{\gamma m-1}[i_0 + 2kp] = 0 \tag{2.36}$$

and

$$v_{\gamma m-1}[i_0] = v_{\gamma m-1}[i_0 + (k-1)p] = \\ = v_{\gamma m-1}[i_0 + (2k-1)p] = 0. \tag{2.37}$$

Next we observe that from the construction of a bi-ideal (see Figure 2.1 and 2.5) and from the equations (2.21) and (2.22) it follows that

$$v_{\gamma m-1}[kp - l - 1, kp] = y[kp - l - 1, kp] = v_{\alpha m-1}1 \tag{2.38}$$

and

$$v_{\gamma m-1}[2kp - l - 2, kp] = y[2kp - l - 2, kp] = v_{\alpha m-1}10. \tag{2.39}$$

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 2.11 — $v_0 \in \text{Suff}\,(v_{\alpha m-1})$), 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 $r \in \{1, 2, \ldots, l-1\}$), then $v_{\gamma m-1}[i_0 + (2k-1)p] = v_{\alpha m-1}[r+1]$. Finally from (2.37) it follows that $v_{\alpha m-1}[r] = v_{\alpha m-1}[r+1] = 0$, i.e., $00 \in \text{F}(y)$, but from lemma 2.16 we know that $00 \notin \text{F}(y)$. This is a contradiction and therefore $z = S_y(x)$ is not ultimately periodic. □

We have shown an example of a universal bi-ideal, i.e., such that for each non-trivial A-sequence the resulting shrunk word is not ultimately periodic. However, the base words of our universal bi-ideal are too short to guarantee good statistical properties of the resulting shrunk word. Nevertheless, we are convinced of the existence of wider class of universal bi-ideals with better statistical properties.

3 Results on bi-ideals

This chapter contains the main results of the thesis. First, in Section 3.1 I present an algorithmic decision procedure for the problem of determining if two bases generate the same finitely generated bi-ideal. Section 3.2 presents some smaller results on finitely-generated bi-ideals — finitely generated bi-ideals are shown to be closed under transformation by morphism and shift operator; a new example showing that finitely generated bi-ideals are not closed under transduction is given, by transforming the bi-ideal generated from (a, b) to the Thue-Morse sequence; finitely generated bi-ideals are shown to be a subclass of morphic sequences. Finally, Section 3.3 generalizes known results on closure properties of ultimately recurrent sequences, by showing that ultimately recurrent sequences are closed under transduction.

3.1 A decision problem in finitely generated bi-ideals

3.1.1 Introduction

It has been known for some time, that there are countably many different finite bases for any finitely generated bi-ideal, yet there was no general algorithm to tell whether two different bases generate the same bi-ideal or not. Some inroads were made by Buls and Lorencs (Buls and Lorencs, 2008), who gave a criterion for determining whether a bi-ideal is periodic by looking at its basis. Later Lorencs (2012) gave a general solution to the problem for the case where the finite basis of a finitely generated bi-ideal consists of only two base words (the basis of the bi-ideal sequence is periodic with period 2).

I solve this problem by giving an efficient decision algorithm. Three kinds of (effective) reductions are presented, that allow to reduce some finite bases of finitely

generated bi-ideals to shorter finite bases (in the sense of the total length of base words) that still generate the same finitely generated bi-ideal (see Definition 3.7). Then it is shown that irreducible bases have a one to one relationships to the finitely generated bi-ideals.

The result includes as a special case the earlier result by Lorencs (2012), and also serves as an alternate criterion to tell whether a bi-ideal is periodic or not. Namely, a finitely generated bi-ideal is periodic if and only if its irreducible basis contains only one base word.

3.1.2 Preliminaries

Lemma 3.1. *If* $(u_0, u_1, \ldots, u_n, \ldots)$ *is a basis of a bi-ideal* x*, then* $(u_0u_1, u_0u_2, \ldots, u_0u_n, \ldots)$ *is also a basis of* x*.*

Proof. Denote by (v_i) the bi-ideal sequence corresponding to the basis $(u_i)_{i\geq 0}$ and by (v'_i) the bi-ideal sequence corresponding to the basis $(u_0u_j)_{j\geq 1}$.

The proof is by induction. Notice $v'_0u_0 = u_0u_1u_0 = v_1$. Assuming by induction $v'_iu_0 = v_{i+1}$

$$v_{i+2} = v_{i+1}u_{i+2}v_{i+1} = v'_iu_0u_{i+2}v'_iu_0 = v'_{i+1}u_0,$$

meaning that $v'_iu_0 = v_{i+1}$ holds for all $i \in \mathbb{N}$. Then $\lim_{i\to\infty} v_i = \lim_{i\to\infty} v'_i$.

□

Corollary 3.2 (Lorencs (2012))**.** *Let* $(u_0, u_1, \ldots, u_n)$ *be a basis of a finitely generated bi-ideal* x*. Then* $(u_0u_1, u_0u_2, u_0u_3, \ldots, u_0u_n, u_0u_0)$ *is also a basis of* x*.*

The corollary follows straightforwardly from lemma 3.1.

We also need some simple results from number theory. Let $\lfloor y \rfloor$ to denote the integer part of a positive rational number y.

Lemma 3.3. *Let* a, b, c *be positive integers. Then*

$$\left\lfloor \frac{a}{b}c \right\rfloor \mod c = \left\lfloor \frac{a \mod b}{b}c \right\rfloor$$

Proof. Note that there exist integers $k \geq 0$ and $0 \leq \ell < b$ such that $a = kb + l$. Then

$$\left\lfloor \frac{a \mod b}{b}c \right\rfloor = \left\lfloor \frac{\ell}{b}c \right\rfloor,$$

while

$$\begin{aligned}\left\lfloor \frac{a}{b}c \right\rfloor \mod c &= \left\lfloor kc + \frac{\ell}{b}c \right\rfloor \mod c \\ &= \left(kc + \left\lfloor \frac{\ell}{b}c \right\rfloor \right) \mod c \\ &= \left\lfloor \frac{\ell}{b}c \right\rfloor \mod c \\ &= \left\lfloor \frac{\ell}{b}c \right\rfloor.\end{aligned}$$

□

Theorem 3.4 (Dirichlet (see, e.g., (Schmidt, 1980))). *Let $\alpha \in \mathbb{R}$ and $N \in \mathbb{N}$, then there are $p, q \in \mathbb{Z}$ such that $1 \leq q \leq N$ and*

$$|q\alpha - p| \leq \frac{1}{N+1}.$$

Corollary 3.5. *Given an integer $k > 1$ and $a, b \in \mathbb{N}$ such that $1 < a < b < ak$, there exist relatively prime $i, j \in \mathbb{N}$ such that*

$$|ia - jb| < \frac{b}{k}$$

with $i \leq j < k$.

Proof. From Theorem 3.4 we can select $p, q \in \mathbb{N}$ such that $1 \leq q < k$ and

$$\left| \frac{a}{b}q - p \right| \leq \frac{1}{k}.$$

Since $b < ak$, then $p \geq 1$. From this we can express

$$\begin{aligned}\left| \frac{a}{b} - \frac{p}{q} \right| &\leq \frac{1}{kq} \\ |aq - pb| &\leq \frac{b}{k}.\end{aligned}$$

Since $a < b$ then it is no problem to select $p \leq q$.

Finally, we divide out the greatest common divisor of p and q to obtain relatively prime i, j such that

$$\frac{i}{j} = \frac{q}{p}.$$

Then

$$|ia - jb| \leq |qa - pb| \leq \frac{b}{k}.$$

□

Lemma 3.6. *Let $a, b, c, n, x, y, z \geq 0$ be integers. Assume also*

$$b = a \mod n + \varepsilon, \tag{3.1}$$

with $|\varepsilon| \leq y$ when

$$x < a \mod n < n - x. \tag{3.2}$$

Then $c - z \leq a \leq c + z$ implies

$$b = c \mod n + \varepsilon', \tag{3.3}$$

with $|\varepsilon'| \leq y + z$ when

$$x + z < c \mod n < n - x - z. \tag{3.4}$$

Proof. We can express $c = kn + l$ and $a = k'n + l'$ for some integers k, k', l, l' with $0 \leq l, l' < n$. Assume (3.4) holds. Then we have $x + z < l < n - x - z$. From $c - z \leq a \leq c + z$ it is clear that $k' = k$ and $x < l' < n - x$, yielding (3.2). Since (3.2) is satisfied, (3.1) holds and we have $b - y \leq l' \leq b + y$. Then it follows from $c - z \leq a \leq c + z$ that $b - y - z \leq l \leq b + y + z$, yielding (3.3). □

3.1.3 Results

It is clear from Corollary 3.2 that any finitely generated bi-ideal has an infinite number of finite bases. We show that it is effectively decidable whether two bases generate the same finitely generated bi-ideal or not. We do this by giving an effective basis reduction algorithm (Definition 3.7) that leads to a finite and unique representation of the finitely generated bi-ideal which we call an irreducible basis (Definition 3.8) of the finitely generated bi-ideal. We prove that there is a one to one correspondence of irreducible bases and finitely generated bi-ideals (Theorem 3.12).

Definition 3.7. *We say a basis $(u_0, u_1, \ldots, u_{n-1})$ 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 naturals k_i such that*

$$u_i = u^{k_i},$$

   for all $i \in \{0, 1, \ldots, n-1\}$. Then (the single element tuple) (u) is also a basis of x;

2. *There is a $T < n$ such that $n = k \cdot T$ for some $k \in \mathbb{N}$ and $u_i = u_{i+T}$ for all $i \in \{1, \ldots, n - T - 1\}$. Then $(u_0, u_1, \ldots, u_{T-1})$ is also a basis of x;*

3. There are such words w_i that

$$u_i = w_{n-1}w_i,$$

for all $i \in \{0, 1, \ldots, n-1\}$. Then $(w_{n-1}, w_0, w_1, \ldots, w_{n-2})$ is also a basis of x.

We use $\{a, b\}$ as the alphabet for examples.

Reduction (1) is just an application of Theorem 1.26. The bi-ideal is a periodic word in this case. An example reduction would be reducing the basis $(ababab, ab, \lambda, abab)$ to just (ab). Both bases generate the bi-ideal $(ab)^\omega$.

Reduction (2) expresses a similarly natural idea. Since a finite basis is just a shorthand for a periodic sequence of base words, if such a basis itself is fully periodic with some period n then the sequence of basis words must also be periodic with this period n. Therefore, it can be represented by a finite basis of length n. For example, $(aab, ab, bb, aab, ab, bb)$ could be reduced to (aab, ab, bb), as both these finite bases still represent the basis sequence $(aab, ab, bb, aab, ab, bb, \ldots)$.

Finally, reduction (3) is the converse of Corollary 3.2. For example, $(aabab, aabbb, aabaab)$ could again be reduced to (aab, ab, bb) and by Corollary 3.2 these would generate the same finitely generated bi-ideal.

Notice that reducing a finite basis always reduces the number and/or total length of its base words. Therefore, we can perform this reduction only a finite number of times before we get to a basis for the bi-ideal that cannot be reduced any further. Also note that each of the conditions necessary for the reductions can be checked for effectively. Finally, the order in which the reductions are applied is unimportant, in the sense that it must always lead to the same irreducible basis (as a corollary of Theorem 3.12).

Definition 3.8. *A finite basis is called* irreducible *if it cannot be further reduced as per Definition 3.7.*

It is natural to split finitely generated bi-ideals into periodic and aperiodic words. Note that we don't have to consider the case of ultimately periodic words, since bi-ideals are recurrent words (and any ultimately periodic and recurrent word is completely periodic). The case of periodic finitely generated bi-ideals was completely solved by Buls and Lorencs (2008) and is reflected in theorem 1.26. What follows before we can state our main result is the analysis of aperiodic bi-ideals. We start with some preliminary results.

Lemma 3.9. *Let x be a bi-ideal generated by a basis (u_i). Then*

$$x = u_{\alpha(1)} u_{\alpha(2)} \cdots u_{\alpha(i)} \cdots ,$$

with $\alpha(i) = \max\{k \mid i \equiv 0 \mod 2^k\}$.

Proof. The proposed decomposition corresponds to the one obtained from definition 1.8 — $x = u_0u_1u_0u_2u_0u_1u_0u_3u_0u_1u_0 \cdots$. Note that in this decomposition every u_k makes its first appearance at position 2^k. From the above formula $\alpha(2^k) = k$.

Note, also, that $\alpha(2^k + l) = \alpha(l)$, when $l < 2^k$. From the structure of x and the previous argument we see that if $\alpha(i)$ is correct for any $i < 2^k$, then it is correct for any $i < 2^{k+1}$. Since it is correct for $i = 0$, it is correct for any i by induction. □

Next we need a slightly more technical lemma. Informally, it states that if a sufficiently long prefix of a finitely generated bi-ideal is periodic, then the whole bi-ideal is periodic.

Lemma 3.10. *Let x be a finitely generated bi-ideal with a basis $(u_0, u_1, \ldots, u_{n-1})$, let $\ell = \max_i |u_i|$. If x has a periodic prefix y with period p and $|y| \geq (2^n + 2)(\ell + p)$, then x is periodic.*

Proof. Let $v_0, v_1, \ldots, v_n, \ldots$ be the bi-ideal sequence associated with $(u_0, u_1, \ldots, u_{n-1})$.

Fix $j = \min\{i \mid p < |v_i|\}$. Either $j = 0$ and $|v_0| = |u_0| \leq \ell$, or

$$v_j = v_{j-1} u_j v_{j-1},$$

in which case $|v_j| \leq 2p + \ell$, because $|v_{j-1}| \leq p$ and $|u_j| \leq \ell$. Either way $|v_j| \leq 2p + \ell$.

By Observation 1.9 x can be expressed as

$$x = v_j \tilde{u}_1 v_j \tilde{u}_2 \cdots v_j \tilde{u}_n \cdots ,$$

where $\tilde{u}_i \in \{u_0, u_1, \ldots, u_{n-1}\}$ for all i. We will show that $v_j u_i v_j$ is periodic with period p for all $i \in \{0, 1, \ldots, n-1\}$. Considering that $|v_j| > p$ it then follows from Corollary 1.4 that x is also periodic with period p.

Let $z = v_{j+n-1} u_j v_j$. Then z is a prefix of v_{j+n} and, therefore, a prefix of x. Moreover, $v_j u_{j+1} v_j, v_{j+1} u_{j+2} v_j, \ldots, v_{j+n-1} u_{j+n} v_j$ are all prefixes of z. Since v_j is a suffix of all v_m with $m \geq j$ then from the previous it follows that $v_j u_i v_j \setminus z$ for all $i \in \{0, 1, \ldots, n-$

1}. To show $v_j u_i v_j$ is periodic with period p for all $i \in \{0, 1, \ldots, n-1\}$ it now suffices to show that $|z| < (2^n + 2)(\ell + p)$ and that z therefore is periodic with period p by the assumptions of the lemma.

$$|z| = |v_{j+n-1} u_j v_j| \leq |v_{j+n-1}| + 2p + 2\ell.$$

Because $|v_k| = |v_{k-1} u_k v_{k-1}| \leq 2|v_{k-1}| + \ell$ then

$$|v_{j+n-1}| \leq 2^{n-1}|v_j| + (2^{n-1} - 1)\ell \leq 2^n p + 2^n \ell - \ell$$

and

$$|z| \leq (2^n + 2)(\ell + p).$$

□

Now we come to what can be considered the main technical result of this section — a property linking two different bases of an aperiodic finitely generated bi-ideal. While the proof of the following Proposition is very involved and technical, it makes the proof of our main result quite straightforward.

Proposition 3.11. *Let $B = (u_0, u_1, \ldots, u_{n-1})$ and $B' = (u'_0, u'_1, \ldots, u'_{n'-1})$ be two finite bases of the same* aperiodic *bi-ideal x and let (v_i) and (v'_i) be the respective bi-ideal sequences associated with B and B'. Then there exist $j, j' \in \mathbb{N}$ such that for all $i \in \mathbb{N}$,*

$$v_{j+i} u_{j+i+1} = v'_{j'+i} u'_{j'+i+1}.^{1}$$

Proof. The proof is loosely structured in 3 parts:

1. choose suitable (large relatively to L) values for j and j';
2. show that the lengths of v_j and $v'_{j'}$ are close;
3. show how this implies the proposition.

Part 1. Denote $L = \max\{|u_1|, |u_2|, \ldots, |u_{n-1}|, |u'_1|, |u'_2|, \ldots, |u'_{n'-1}|\}$ and $N = \max\{n, n'\}$. It is possible to select j and j' such that

$$|v_j| > 100\left(2^{N+1} + 4\right)^2 L, \tag{3.5}$$

$$|v'_{j'}| > 100\left(2^{N+1} + 4\right)^2 L, \tag{3.6}$$

$$1 \geq \frac{|v'_{j'}|}{|v_j|} > \frac{7}{10}. \tag{3.7}$$

[1] Note that u_{j+i+1} in the above represents $u_{j+i+1 \mod n}$ while $u'_{j'+i+1}$ represents $u'_{j+i+1 \mod n'}$.

Note that $L \leq 0.001|v'_{j'}| \leq 0.001|v_j|$, because $N, L \geq 1$. While it is easy to satisfy (3.5) and (3.6) by simply choosing j and j' large enough, we need to show that (3.7) can also be satisfied.

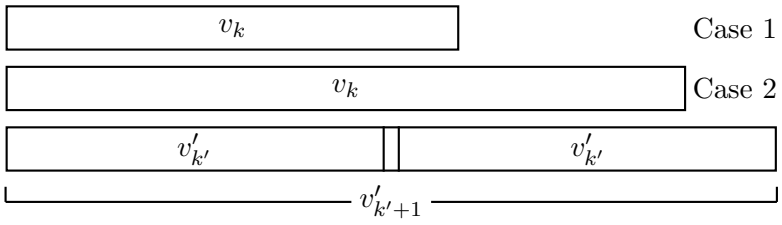

Figure 3.1: Selecting j and j' so that equation (3.7) is satisfied.

We start off by selecting k and k' such that (3.5) and (3.6) are satisfied when $j = j' = \min\{k, k'\}$ and also

$$|v'_{k'}| \leq |v_k| \leq |v'_{k'+1}|.$$

If (3.7) is satisfied by taking $j = k$ and $j' = k'$, we are done (Case 1 in Figure 3.1). If it is not, we can swap B with B' (without loosing generality) and take $j = k' + 1$ and $j' = k$ (Case 2 in Figure 3.1). To see that these j and j' indeed satisfy (3.7) note that $|v'_{k'+1}| \leq 2|v'_{k'}| + L \leq 2.001|v'_{k'}|$. From

$$\frac{|v'_{k'}|}{|v_k|} \leq \frac{7}{10},$$

it follows

$$\frac{|v_k|}{|v'_{k'+1}|} \geq \frac{|v_k|}{2.001|v'_{k'}|} \geq \frac{10}{2.001 * 7} > \frac{7}{10}.$$

For the remainder of the proof of this Proposition, we shall denote $v_j = v$ and $v'_{j'} = v'$.

Part 2. Assume

$$\frac{|v'|}{|v|} < \frac{2^{N+1} + 3}{2^{N+1} + 4}. \tag{3.8}$$

We shall argue for contradiction by showing how this implies x to be periodic.

From Observation 1.9 we know we can express

$$x = v\tilde{u}_0 v\tilde{u}_1 v \cdots v\tilde{u}_n \cdots$$
$$x = v'\tilde{u}'_0 v'\tilde{u}'_1 v' \cdots v'\tilde{u}'_n \cdots,$$

where $\tilde{u}_i \in \{u_0, \ldots, u_{n-1}\}$ and $\tilde{u}'_i \in \{u'_0, \ldots, u'_{n'-1}\}$ for all i. For $i \geq 0$ denote

$$w_i = \prod_{k=0}^{i} v\tilde{u}_k$$
$$w'_i = \prod_{k=0}^{i} v'\tilde{u}'_k.$$

What follows is by far the most involved part of the proof. Our ultimate goal is to show that x is periodic when assumption (3.8) holds. To do this, we exploit the offsets of occurrences of v's relative to occurrences of v''s (see Figure 3.2). Before we can do that we need to establish some technique to deal with the uncertainties introduced by the u_i's and u'_i's.

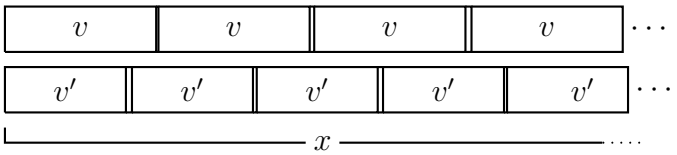

Figure 3.2: Decompositions of x into v's and v''s.

We start off by showing that there exist relatively prime integers $1 < T \leq T' < 2(2^{N+1} + 4)$ such that for all $i \geq 1$

$$w_{iT-1} = w'_{iT'-1}. \tag{3.9}$$

It is easy to see that $T, T' > 1$. From (3.8) and (3.5) we have $|v| - |v'| > |v|/(2^{N+1}+4) > L$ and, therefore, $|w_0| > |w'_0|$ and it cannot be that $T = T' = 1$. If only one of T, T' were assumed to be one (let it be T), then

$$T'|v'| - L \leq |v| \leq T'(|v'| + L),$$

implying $|v'| < 0.7|v|$ and contradicting (3.7). So $T, T' > 1$.

By Corollary 3.5 we can select relatively prime integers $T \leq T' < 2(2^{N+1}+4)$ such that

$$\big|T|v| - T'|v'|\big| \leq \frac{|v|}{2(2^{N+1}+4)}.$$

For these

$$T|v| \leq |w_{T-1}| \leq T(|v| + L),$$
$$T'|v'| \leq |w'_{T'-1}| \leq T'(|v'| + L),$$

therefore (taking into account that $TL \leq T'L$),

$$\begin{aligned} \big||w_{T-1}| - |w'_{T'-1}|\big| &\leq \big|T|v| - T'|v'|\big| + T'L \\ &\leq \frac{|v|}{2(2^{N+1}+4)} + \frac{T'|v|}{100(2^{N+1}+4)^2} \\ &\leq \frac{|v|}{2(2^{N+1}+4)} + \frac{|v|}{50(2^{N+1}+4)} \\ &\leq \frac{13|v|}{25(2^{N+1}+4)}, \end{aligned} \tag{3.10}$$

and since $|v'| > 0.7|v|$ (from Equation (3.7)),

$$\frac{13|v|}{25(2^{N+1}+4)} < \frac{26|v'|}{35(2^{N+1}+4)} < \frac{|v'|}{2^{N+1}+4}.$$

Assume now, that $w_{T-1} \neq w'_{T'-1}$. Then $w_{T-1} = w'_{T'-1}s$ (or in the symmetrical case $w_{T-1}s = w'_{T'-1}$) for some word s with

$$|s| = \big||w_{T-1}| - |w'_{T'-1}|\big| < \frac{|v'|}{2^{N+1}+4}.$$

Because both $w'_{T'-1}v' \in \operatorname{Pref} x$ and $w_{T-1}v' \in \operatorname{Pref} x$ (because $v' \in \operatorname{Pref} v$) there must exist words r and t such that $v' = sr = rt$ (see Figure 3.3). This means v' must be periodic with period $|s|$ (by Theorem 1.2).

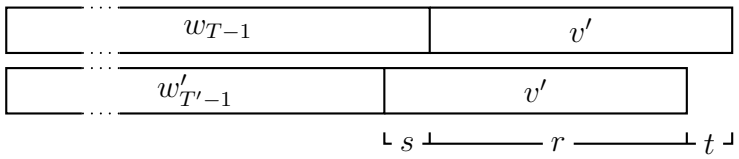

Figure 3.3: Decomposition $v' = sr = rt$.

We can now apply Lemma 3.10 to $v' \in \operatorname{Pref} x$. From the lemma it follows that if $|v'| \geq (2^N+2)(L+|s|)$ then x is periodic. Indeed,

$$\begin{aligned} (2^N+2)(L+|s|) &\leq (2^N+2)\left(\frac{|v'|}{100(2^{N+1}+4)^2} + \frac{|v'|}{2^{N+1}+4}\right) \\ &< (2^N+2)\left(\frac{1}{2^{N+1}+4} + \frac{1}{2^{N+1}+4}\right)|v'| = |v'|. \end{aligned}$$

This would contradict the assumption that x is aperiodic and, therefore, $w_{T-1} = w'_{T'-1}$.

The rest of the proof that $w_{iT-1} = w'_{iT'-1}$ is by induction on i.

$$T|v| \leq |w_{iT-1}| - |w_{(i-1)T-1}| \leq T(|v|+L)$$
$$T'|v'| \leq |w'_{iT'-1}| - |w'_{(i-1)T'-1}| \leq T'(|v'|+L).$$

Note, that if $w_{(i-1)T-1} = w'_{(i-1)T'-1}$ then the previous implies

$$\big||w_{iT-1}| - |w'_{iT'-1}|\big| \leq \big|T|v| - T'|v'|\big| + T'L$$

which is the same as the first valuation in Equation (3.10). The rest of the proof proceeds analogously to the rest of the proof of the case $w_{T-1} = w'_{T'-1}$.

Next we show that v' and v are periodic with period $P < 0.5|v'|$. Indeed, consider the relative positions of the last v' and the last v in $w'_{T'-1} = w_{T-1}$. From $v' \in \operatorname{Pref} v$, and equations (3.7) and (3.8) it follows that there must exist words s, t and r such that $v' = sr = rt$ (Figure 3.4), moreover, $0 < |s| < 0.3|v| + L < 0.5|v'|$. This implies that v' is periodic with period $P = |s|$ (by Theorem 1.2). To see that v is periodic with period P too, note that it is mostly covered by srt and srt is periodic with period P by Corollary 1.4. If v is not fully covered by srt, one just has to look at the previous overlap of v and v' (at the end of w_{T-2} and $w'_{T'-2}$) which covers the small suffix (of length not exceeding L) not covered by srt.

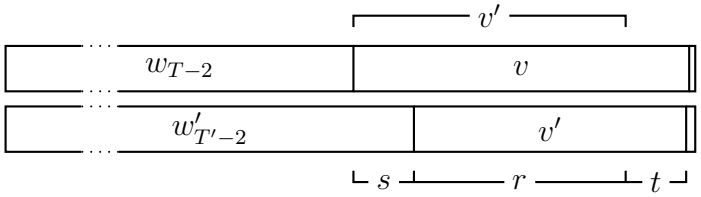

Figure 3.4: The last v and v' in $w'_{T'-1} = w_{T-1}$.

To show that x is periodic and arrive at our contradiction, we will now prove that that vu_iv is also periodic with the same period P for all i.

We start by finding an estimate for

$$\xi(i) = |w_i| - |w'_{\gamma(i)}|, \tag{3.11}$$

where

$$\gamma(i) = \max\{i' \mid |w'_{i'}| \leq |w_i|\}. \tag{3.12}$$

For this context we define $w_{-1} = \lambda$. See Figure 3.5 for a visualisation. Note that $\xi(i)$ is a number and not a word.

Since T and T' are relatively prime, at least one of them must be odd. We can assume without loss of generality this to be T. If we make this assumption, we must make sure not rely on $T \leq T'$ or on $|v| > |v'|$.

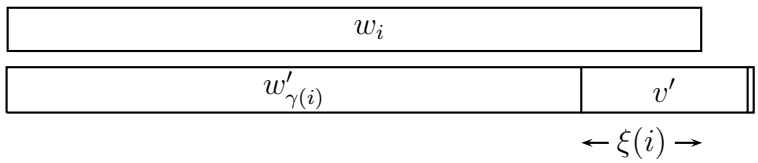

Figure 3.5: $\xi(i)$ and $\gamma(i)$.

We start by defining $\hat{w}_i = (v')^{i+1}$ and

$$\hat{\gamma}(i) = \max\{i' \mid |\hat{w}_{i'}| \leq |w_i|\}$$
$$\hat{\xi}(i) = |w_i| - |\hat{w}_{\hat{\gamma}(i)}|.$$

By these definitions $|\hat{w}_{\hat{\gamma}(i)}| = (\hat{\gamma}(i)+1)|v'|$ and

$$\hat{\xi}(i) = |w_i| \mod |v'|.$$

We use this as a starting point.

For now assume $i < T$. We will be able to drop this assumption later, but it greatly simplifies the derivations early on.

Note that $\gamma(i) < T'$, because $\gamma(T-1) = T'-1$ from $w_{T-1} = w'_{T'-1}$. A simple fact we will frequently use is that $T'L, TL < 0.005|v'| - 1$. This follows from (3.5) and (3.6) and the fact that $N \geq 1$ by definition. Then,

$$T'L \leq \frac{2(2^{N+1}+4)|v'|}{100(2^{N+1}+4)^2} < 0.003|v'| - 1,$$
$$TL \leq \frac{2(2^{N+1}+4)|v|}{100(2^{N+1}+4)^2} < 0.003|v| - 1,$$

and $TL < 0.005|v'| - 1$ follows from (3.7).

We can estimate,

$$0 \leq |w'_{\gamma(i)}| - |\hat{w}_{\gamma(i)}| \leq \gamma(i)L \leq T'L < 0.005|v'|,$$

meaning that it could be the case that $\gamma(i) = \hat{\gamma}(i) - 1$, but only when $\hat{\xi}(i) < 0.005|v'|$, in which case $\xi(i) > 0.995|v'|$ (illustrated in Figure 3.6).

So we have two cases - either $\gamma(i) = \hat{\gamma}(i)$ and

$$\hat{\xi}(i) - 0.005|v'| < \xi(i) \leq \hat{\xi}(i), \tag{3.13}$$

or $\gamma(i) = \hat{\gamma}(i) - 1$ and

$$0.995|v'| \leq \xi(i) < |v'| + L,$$

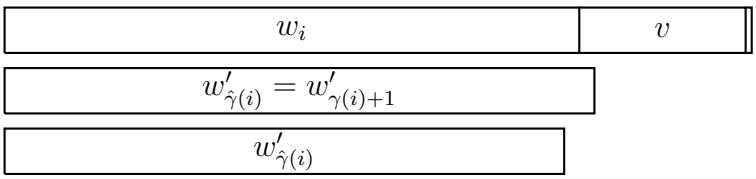

Figure 3.6: The case when $\gamma(i) \neq \hat{\gamma}(i)$.

while $\hat{\xi}(i) \leq 0.005|v'|$. Since we will not really be interested in the case when $\xi(i) \geq 0.995|v'|$, we will consider only the case when $\hat{\xi}(i) > 0.005|v'|$ and use Equation (3.13) to express $\xi(i)$ in terms of $\hat{\xi}(i)$ and an integer error term $|\varepsilon_i| \leq 0.005|v'|$ to obtain

$$\xi(i) = \hat{\xi}(i) + \varepsilon_i, \tag{3.14}$$

when $\hat{\xi}(i) > 0.005|v'|$.

Let

$$\tilde{\xi}(i) = (i+1)|v| \mod |v'|. \tag{3.15}$$

Since

$$(i+1)|v| \leq |w_i| \leq (i+1)(|v| + L),$$

and $(i+1)L < 0.005|v'|$ from $i < T$, we can use Lemma 3.6 to update Equation (3.14) to become

$$\xi(i) = \tilde{\xi}(i) + \tilde{\varepsilon}_i, \tag{3.16}$$

with $|\tilde{\varepsilon}_i| \leq 0.010|v'|$ when $0.010|v'| < \tilde{\xi}(i) < 0.990|v'|$.

By now we have confined any uncertainty caused by the base words into the error term. Our next step is to do the same with the uncertainty about $|v|/|v'|$. Recall that

$$T|v| \leq |w_{T-1}| \leq T(|v| + L) < T|v| + 0.005|v'| - 1$$
$$T'|v'| \leq |w'_{T'-1}| \leq T'(|v'| + L) < T'|v'| + 0.005|v'| - 1.$$

Since $|w_{T-1}| = |w'_{T'-1}|$ the previous implies

$$T'|v'| - 0.005|v'| + 1 \leq T|v| \leq T'|v'| + 0.005|v'| - 1,$$

or expressed with an error term $|\varepsilon| < 0.005|v'| - 1$

$$|v| = \frac{T'}{T}|v'| + \frac{\varepsilon}{T}.$$

Combining this with Equation (3.15), we define

$$\tilde{\tilde{\xi}}(i) = \left\lfloor \frac{(i+1)T'}{T}|v'| \right\rfloor \mod |v'|. \tag{3.17}$$

Since

$$\left|(i+1)|v| - \left\lfloor \frac{(i+1)T'}{T}|v'| \right\rfloor\right| \leq \left|\left\lfloor \frac{(i+1)\varepsilon}{T} \right\rfloor\right| + 1 < 0.005|v'|,$$

we can use Lemma 3.6 to update (3.16) to

$$\xi(i) = \tilde{\tilde{\xi}}(i) + \tilde{\tilde{\varepsilon}}_i, \tag{3.18}$$

with $|\tilde{\tilde{\varepsilon}}_i| < 0.015|v'|$ when $0.015|v'| < \tilde{\tilde{\xi}}(i) < 0.985|v'|$. Using Lemma 3.3 we can rewrite (3.17) as

$$\tilde{\tilde{\xi}}(i) = \left\lfloor \frac{(i+1)T' \mod T}{T}|v'| \right\rfloor. \tag{3.19}$$

Until now we had assumed $i < T$. Now we look at $\xi(kT+i)$, where $i < T$ and $k > 0$. From

$$|w_{kT-1}| + |w_i| - (i+1)L \leq |w_{kT+i}| \leq |w_{kT-1}| + |w_i| + (i+1)L,$$
$$|w_{kT'-1}| + |w'_{\gamma(i)}| - (\gamma(i)+1)L \leq |w'_{kT'+\gamma(i)}| \leq |w'_{kT'-1}| + |w'_{\gamma(i)}| + (\gamma(i)+1)L,$$

we have

$$\xi(i) - 0.010|v'| \leq \left||w_{kT+i}| - |w'_{kT'+\gamma(i)}|\right| \leq \xi(i) + 0.010|v'|.$$

From this it follows that $\gamma(kT+i) = kT' + \gamma(i)$ at least for the case when $0.010|v'| < \xi(i) < 0.990|v'|$ (because, $0 \leq \xi(kT+i) < |v'| + L$ by definition) and that in this case

$$\xi(kT+i) = \xi(i) + \varepsilon_{k,i}, \tag{3.20}$$

for some error term $|\varepsilon_{k,i}| \leq 0.010|v'|$.

From Equation (3.19) it is clear that $\tilde{\tilde{\xi}}(kT+i) = \tilde{\tilde{\xi}}(i)$ and using Lemma 3.6 it follows from (3.20) that Equation (3.18) holds with $|\tilde{\tilde{\varepsilon}}_i| \leq 0.025|v'|$ for all $i \in \mathbb{N}$ when $0.025|v'| < \tilde{\tilde{\xi}}(i) < 0.975|v'|$.

We can now proceed to show that x is periodic. We will show that vu_iv is periodic for all $i \in \{0, 1, \ldots, n-1\}$ by showing how overlaps of vu_iv by v' induce this periodicity. Recall that T could be assumed odd, and that T' and T are relatively prime. Then it follows from Equations (3.18) and (3.19) that we can select i_1 and i_2 such that for all $k \in \mathbb{N}$

$$0.25|v'| + L < \xi(i_1 + kT) < 0.5|v'| - L \tag{3.21}$$
$$0.5|v'| + L < \xi(i_2 + kT) < 0.75|v'| - L, \tag{3.22}$$

because $(i+1)T'$ form a complete residue system modulo T.

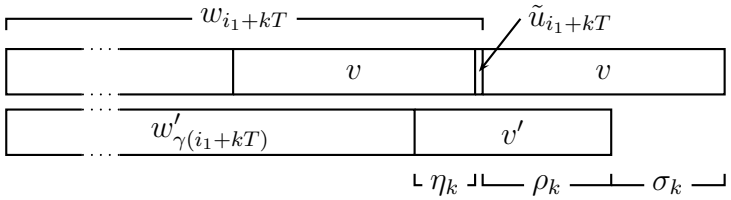

Figure 3.7: Overlap of v and v' at w_{i_1+kT}.

Now we can express (see Figure 3.7)

$$w_{i_1+kT}v = w'_{\gamma(i_1+kT)}\eta_k\tilde{u}_{i_1+kT}\rho_k\sigma_k,$$

in such a way that

$$|\eta_k| + |\tilde{u}_{i_1+kT}| = \xi(i_1 + kT)$$
$$v' = \eta_k\tilde{u}_{i_1+kT}\rho_k$$
$$v = \rho_k\sigma_k.$$

Notice that $\xi(i_1 + kT) < 0.5|v'| - L$ implies $|\rho_k| > 0.5|v'| > P$, therefore $\eta_k\tilde{u}_{i_1+kT}v$ is periodic with period P. Since $|\eta_k| > 0.25|v'|$ (from Equation (3.21)) and $\eta_k \in \text{Suff } v$ for all $k \in \mathbb{N}$, we can choose a maximal $\eta \in \text{Suff } \eta_k$ for all $k \in \mathbb{N}$ such that $|\eta| > 0.25|v'| > 0.5P$. Then $\eta\tilde{u}_{i_1+kT}v$ is periodic with period P for all $k \in \mathbb{N}$.

Very similarly we can express (see Figure 3.8)

$$w_{i_2+kT}y_k = w'_{\gamma(i_2+kT)}x_k\tilde{u}_{i_2+kT}y_k,$$

so that

$$|x_k| + |\tilde{u}_{i_2+kT}| = \xi(i_2 + kT),$$
$$v' = x_k\tilde{u}_{i_2+kT}y_k$$

for all $k \in \mathbb{N}$. Analogously to the above, it follows from Equation (3.22) that $|x_k| > 0.5|v'| > P$, implying $v\tilde{u}_{i_2+kT}y_k$ is periodic for all $k \in \mathbb{N}$. Since $|x_k| < 0.75|v'| - L$ we can choose a y (because $y_k \in \text{Pref } v$) with $|y| > 0.25|v'| > 0.5P$ such that $v\tilde{u}_{i_2+kT}y$ is periodic with period P for all $k \in \mathbb{N}$.

What remains to show is that there are $l_0, l_1, \ldots, l_{n-1}$ and $m_0, m_1, \ldots, m_{n-1}$ such that $\tilde{u}_{i_1+l_iT} = u_i$ and $\tilde{u}_{i_2+m_iT} = u_i$. Then we would have shown $\eta u_i v$ and $v u_i y$ periodic

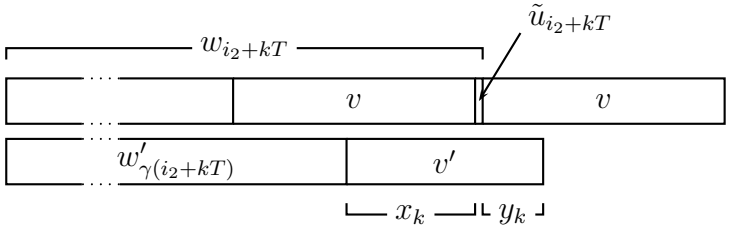

Figure 3.8: Overlap of v and v' at w_{i_2+kT}.

with period $P < 0.5|v'|$ for all $i \in \{0, 1, \ldots, n-1\}$. Since $|\eta| > 0.5P$, $|y| > 0.5P$ and $\eta u_i y \smallsetminus v u_i v$, this would imply that $v u_i v$ is periodic with period P for all $i \in \{0, 1, \ldots, n-1\}$ and that therefore x is periodic, giving our contradiction.

To see that there are such l_i and m_i, notice that we can express

$$\tilde{u}_{i_1+kT} = u_{\alpha((i_1+kT+1)2^{j+1})} \mod n$$
$$\tilde{u}_{i_2+kT} = u_{\alpha((i_2+kT+1)2^{j+1})} \mod n$$

using function α from Lemma 3.9. Since T is odd, it is clear from the definition of α that both $\alpha((i_1 + kT + 1)2^{j+1})$ and $\alpha((i_2 + kT + 1)2^{j+1})$ eventually assume every integer value greater than or equal to $j + 1$, so that their values modulo n assume every value in $\{0, 1, \ldots, n-1\}$. This concludes the proof that

$$\frac{|v'|}{|v|} \geq \frac{2^{N+1} + 3}{2^{N+1} + 4}. \tag{3.23}$$

Part 3. In this final part of the proof we show how Equation (3.23) and Lemma 3.10 implies

$$w_i = w'_i,$$

for all $i \in \mathbb{N}$ and how this implies the statements of the proposition.

The proof of $w_i = w'_i$ is by induction on i and is virtually identical to the proof of Equation (3.9). Assume by induction $w_{i-1} = w'_{i-1}$ (recall that $w_{-1} = w'_{-1} = \lambda$). Assume for the sake of contradiction, that $w_i \neq w'_i$. Then $w_i s = w'_i$ for some word s (or in the symmetric case $w_i = w'_i s$). By Equation (3.23)

$$|s| \leq \frac{|v|}{2^{N+1} + 4} + L,$$

and there exist words r and t such that $v' = sr = rt$ (the reasoning is the same as illustrated in Figure 3.3 for the proof of (3.9)). Then v' must be periodic with period $|s|$

by Theorem 1.2. Since $v' \in \operatorname{Pref} x$ and v' is periodic, it follows from Lemma 3.10 that x must be periodic because (considering $|v'| \leq 8/7|v|$ from (3.23) because $N \geq 1$)

$$(2^N + 2)(L + |s|) \leq (2^N + 2)\left(\frac{2|v'|}{100(2^{N+1}+4)^2} + \frac{8|v'|}{7(2^{N+1}+4)}\right) < |v'|.$$

This would be a contradiction to the assumption of x being aperiodic and, therefore, $w_i = w'_i$. This directly implies the proposition, because

$$v_{j+i}u_{j+i+1} = w_{2^i-1} = w'_{2^i-1} = v'_{j'+i}u'_{j'+i+1}.$$

□

Now we are ready to prove our main result.

Theorem 3.12. *There is one and only one irreducible basis for any finitely generated bi-ideal.*

Proof. The case when the bi-ideal is a periodic word is resolved by Theorem 1.26 which implies that a bi-ideal is periodic if and only if its irreducible basis contains exactly one base word. Moreover, this word has to be the shortest period of the periodic word because otherwise it could be further reduced using reduction 1. Since the shortest period uniquely describes any periodic word, an irreducible basis provides a unique representation of a periodic bi-ideal.

Next we have to consider the case when the bi-ideal is aperiodic. Assume that there are two different irreducible bases $(u_0, u_1, \ldots, u_{n-1})$ and $(u'_0, u'_1, \ldots, u'_{n'-1})$ that generate the same aperiodic bi-ideal x. We argue for contradiction by showing that at least one of the bases can be reduced further.

We first show $n = n'$. Let $(v)_i$ and $(v')_i$ be the bi-ideal sequences associated with the bases $(u)_i$ and $(u')_i$, respectively. From Proposition 3.11 we know that there exist j and j' such that for all $i \in \mathbb{N}$

$$v_{j+i}u_{j+i} = v'_{j'+i}u_{j'+i}. \tag{3.24}$$

Without loss of generality we can assume $v_j \geq v'_{j'}$. Then $v_j = v'_{j'}\nu$ for some word ν. Since

$$\begin{aligned} |v_{j+i+1}| - |v'_{j'+i+1}| &= |v_{j+i}u_{j+i+1}| - |v'_{j'+i}u'_{j'+i+1}| + |v_{j+i}| - |v'_{j'+i}| \\ &= |v_{j+i}| - |v'_{j'+i}|, \end{aligned}$$

by induction $v_{j+i} = v'_{j'+i}\nu$ and therefore,

$$u'_{j'+i} = \nu u_{j+i} \tag{3.25}$$

for all $i \in \mathbb{N}$. Note that (3.25) implies that both bases $(u)_i$ and $(u')_i$ are periodic with period $\gcd(n, n')$. Unless $n = n'$ this implies that at least one of the bases can be reduced using reduction 2. This would contradict the assumption that the bases are irreducible, so $n = n'$.

Since the base lengths are equal, the only way in which the bases can be different is when either $|\nu| > 0$ or when the basis words are rotated, namely, $u_0 = u'_k$ for some $k \in \{1, \ldots, n-1\}$. The second option can be dismissed, because then the basis had to be periodic with period $\gcd(n, k)$ and could be reduced using reduction 2. So we have to assume $|\nu| > 0$.

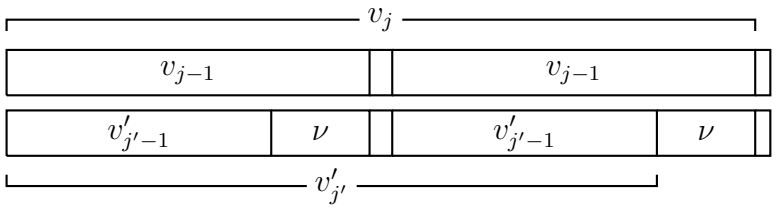

Figure 3.9: v_{j-1} and $v'_{j'-1}$.

We want to show that there is a k such that $v_k = \nu$. Let's look at $v_j = v'_{j'}\nu$.

If we assume $j = 0$, then $v_0 = u_0 = v'_{j'}\nu$, meaning that u_0 is not a factor of $v'_{j'}$. Since $u'_{j'+i} = \nu u_{j+i}$ for all i, the previous means $0 < j' < n$ and $u_0 = u_{n-j'}$. Then the bases have to be periodic with period $\gcd(n, n - j') < n$ meaning that they could be reduced using reduction 2. So we have a contradiction, and $j > 0$.

If we assume $j' = 0$, then

$$v_j u_{j+1} = v'_0 u'_1 = v'_0 \nu u_{j+1} = \nu u_j \nu u_{j+1},$$

meaning that $v_{j-1} = \nu$, giving $k = j - 1$. Finally, if we assume both $j, j' > 0$, then we can decompose $v'_{j'} = v'_{j'-1} u'_j v'_{j'-1}$ and $v_j = v_{j-1} u_j v_{j-1}$. Since $u'_{j'} = \nu u_j$, we find that $v_{j-1} = v'_{j'-1}\nu$ and that (3.24) holds for v_{j-1} and $v'_{j'-1}$ as well (see Figure 3.9) and we can continue to apply these same arguments to $j - i$ and $j' - i$ backwards until we reach a situation where $j' - i = 0$ and $v_k = \nu$ with $k = j - i - 1$.

The previous implies $u'_i = v_k u_{k+i+1}$ for all $i \in \{0, 1, \ldots, n-1\}$. If $k = 0$, then $u'_i = u_0 u_{i+1}$ and can be reduced using reduction 3 to $u''_i = u_i$. If $k > 0$, we can express $v_k = v_{k-1} u_k v_{k-1}$ and then

$$u'_i = v_{k-1} u_k v_{k-1} u_{k+i+1},$$

and can be reduced using reduction 3 to become

$$u_i'' = v_{k-1}u_{k+i}.$$

In either case this is a contradiction to the assumption that both bases are irreducible. Therefore $|\nu| = 0$ and the bases are equal. □

3.2 Some closure properties of finitely generated bi-ideals

3.2.1 Closure under morphism and left shift

Theorem 3.13. *Let μ be a morphism and x a finitely generated bi-ideal. If $\mu(x)$ is not the empty word, then $\mu(x)$ is a finitely generated bi-ideal.*

Proof. The proof is almost obvious. Since the basis (u_i) of x is periodic, so will be the sequence $(\mu(u_i))_{i\geq 0}$. If $\mu(x)$ is not empty, there must be some first j such that $\mu(u_j)$ is not empty. If we look at the sequence $(\mu(u_i))_{i\geq j}$ we see that this will be a valid periodic basis for $\mu(x)$. Therefore, $\mu(x)$ is a finitely generated bi-ideal. □

Definition 3.14. *We call the operation $S(x)$ a* left shift *of x if there is a letter a and a word y such that $x = ay$ and $S(x) = y$.*

Theorem 3.15. *Let x be a finitely generated bi-ideal. Then $S(x)$ is also a finitely generated bi-ideal.*

Proof. Let $(u_0, u_1, \ldots, u_{n-1})$ be a finite basis of x. We can assume without loss of generality that $|u_0| > 1$ (such a basis can always be produced from a basis with $|u_0| = 1$ by applying Corollary 3.2) and (v_i) be the associated bi-ideal sequence.

We construct a basis (u_i') for $S(x)$ as follows. Take $u_0' = S(u_0)$ and $u_i' = u_i u_0[0]$ (recall that $u_0[0]$ denotes the first letter of u_0) for all $i > 0$. Let (v_i') be the associated bi-ideal sequence of (u_i'). Then

$$v_i' = S(v_i),$$

and (u_i') is a basis of $S(x)$.

(u_i') is only ultimately periodic, however. If we apply to it the transformation of Lemma 3.1 we get a periodic basis for $S(x)$ and so $S(x)$ is a finitely generated bi-ideal. □

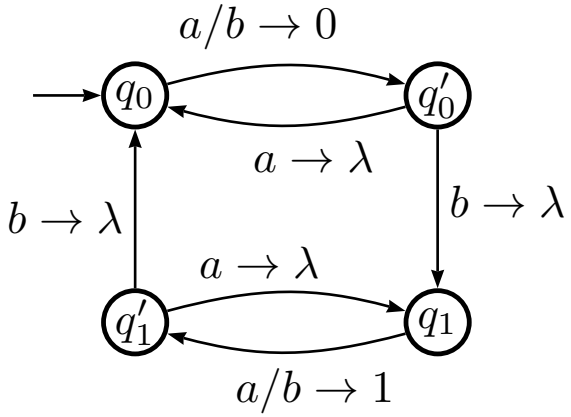

Figure 3.10: The transducer that transforms the finitely-generated bi-ideal x with basis (a, b) into the Thue-Morse word.

3.2.2 Transducing a finitely generated bi-ideal into the Thue-Morse word

Definition 3.16. *Let ρ be an inverting coding on $\{0, 1\}$ (t.i., $\rho(0) = 1$ and $\rho(1) = 0$). Define a sequence of words (t_i) such that $t_0 = 0$ and*

$$t_i = t_{i-1}\rho(t_{i-1}),$$

for all $i > 0$. Then the ω-word

$$\tau = \lim_{n\to\infty} t_n$$

is called the Thue-Morse *word.*

Theorem 3.17. *There is a transducer that transforms a finitely generated bi-ideal into the Thue-Morse word.*

Proof. We will prove the theorem by constructing a transducer $T = \langle Q, \{a, b\}, \{0, 1\}, q_0, \circ, * \rangle$ that transduces the finitely generated bi-ideal x with basis (a, b) into the Thue-Morse word τ.

Let $Q = \{q_0, q_0', q_1, q_1'\}$ and define $\circ$ and $*$ to be

$$\begin{array}{cccc} q_0 \circ a = q_0' & q_0 \circ b = q_0' & q_0' \circ a = q_0 & q_0' \circ b = q_1 \\ q_1 \circ a = q_1' & q_1 \circ b = q_1' & q_1' \circ a = q_1 & q_1' \circ b = q_0 \\ q_0 * a = 0 & q_0 * b = 0 & q_0' * a = \lambda & q_0' * b = \lambda \\ q_1 * a = 1 & q_1 * b = 1 & q_1' * a = \lambda & q_1' * b = \lambda. \end{array}$$

The graph corresponding to T is shown in Figure 3.10.

Notice that because of the symmetry of T

$$q_0 * u = \rho(q_1 * u),$$

where ρ is the inverting coding $\rho(0) = 1$ and $\rho(1) = 0$.

Let (v_i) be the bi-ideal sequence associated with the basis (a, b) and t_i the sequence of words that produces the Thue-Morse word as per Definition 3.16. We will show that $T(v_i) = t_i$ by induction. Notice, first, that $T(v_0) = T(a) = 0 = t_0$. Assume, next, that $T(v_{i-1}) = t_{i-1}$. We will distinguish two cases.

Case 1 — i is odd. If i is odd, the last letter of t_{i-1} is a 0. Since $T(v_{i-1}) = q_0 * v_{i-1} = t_{i-1}$, it must be that $q_0 \circ v_{i-1} = q_0'$. Since $v_i = v_{i-1}u_iv_{i-1}$, and because $u_i = b$ when i is odd,

$$q_0 * v_i = (q_0 * v_{i-1})\,((q_0' \circ b) * v_{i-1}) = (q_0 * v_{i-1})(q_1 * v_{i-1}) = t_{i-1}\rho(t_{i-1}) = t_i.$$

Case 2 — i is even. When i is even, the last letter of t_{i-1} is a 1 and, therefore, $q_0 \circ v_{i-1} = q_1'$. However, $u_i = a$ for even i. And so, for even i

$$q_0 * v_i = (q_0 * v_{i-1})\,((q_1' \circ a) * v_{i-1}) = (q_0 * v_{i-1})(q_1 * v_{i-1}) = t_{i-1}\rho(t_{i-1}) = t_i.$$

This concludes the proof of the inductive step. □

3.2.3 Finitely generated bi-ideals and morphic sequences

Theorem 3.18. *Any finitely generated bi-ideal is a morphic sequence.*

Proof. Let $R_a : \Sigma^* \to (\Sigma \setminus \{a\})^*$ denote the operation of removing the character a from a word. For example, $R_b(aabacbaac) = aaacaac$.

Suppose $x \in \Sigma^\omega$ is a bi-ideal generated by $(u_0, u_1, \ldots, u_{n-1})$ and (v_i) is the bi-ideal sequence associated with this basis. Then we can define a morphism $\mu : \Sigma \bigcup \{\varkappa\} \to \Sigma \bigcup \{\varkappa\}$ (where $\varkappa \notin \Sigma$) such that μ is the identity on the elements of Σ while

$$\mu(\varkappa) = \varkappa u_0 \varkappa u_1 \varkappa u_0 \varkappa u_2 \cdots \varkappa u_{n-1} \varkappa \cdots \varkappa u_0 \varkappa u_1 \varkappa u_0,$$

i.e. we define μ in such a way that $R_\varkappa(\mu(\varkappa)) = v_{n-1}$.

We proceed to show by induction, that $R_\varkappa(\mu^k(\varkappa)) = v_{kn-1}$. Assume by induction $R_\varkappa(\mu^{k-1}(\varkappa)) = v_{(k-1)n-1}$. Because

$$v_{(k-1)n-1} = v_{(k-2)n-1}u_0v_{(k-2)n-1}u_1v_{(k-2)n-1}u_0v_{(k-2)n-1} \cdots u_n \cdots u_0v_{(k-2)n-1},$$

it must be that

$$\mu^{k-1}(\varkappa) = \mu^{k-2}(\varkappa)u_0\mu^{k-2}(\varkappa)u_1\mu^{k-2}(\varkappa)u_0 \cdots u_n \cdots \mu^{k-2}(\varkappa)u_0\mu^{k-2}(\varkappa),$$

and, therefore,

$$\mu^{k}(\varkappa) = \mu^{k-1}(\varkappa)u_0\mu^{k-1}(\varkappa)u_1\mu^{k-1}(\varkappa)u_0 \cdots u_n \cdots \mu^{k-1}(\varkappa)u_0\mu^{k-1}(\varkappa),$$

meaning that

$$R_\varkappa(\mu^k(\varkappa)) = v_{kn-1}.$$

Now we can see that $R_\varkappa(\mu^\omega(\varkappa)) = x$. Since $R_\varkappa$ can be easily implemented by a transducer and because morphic words are closed under transformation by a transducer (Dekking, 1994), x is a morphic word.

□

3.3 A closure property of ultimately recurrent sequences

In this section we extend a result by Buls (2005) who showed that ultimately recurrent (bi-ideal) words (see Definition 1.5) are always transformed to ultimately recurrent words by Mealy machines. We show that the same holds when they are transformed by a transducer.

Transducer-invariance has been studied before and has been shown for ultimately uniformly recurrent sequences and some of their generalizations in (Muchnik et al., 2003), for morphic sequences in (Dekking, 1994) and for primitive morphic sequences in (Holton and Zamboni, 2000).

We will denote by $\mathcal{R}$ the class of all recurrent sequences, and by $\mathcal{UR}$ the class of all ultimately recurrent sequences. Obviously $\mathcal{R} \subset \mathcal{UR}$.

We show, that ultimately recurrent infinite sequences are invariant under transformation by a transducer (provided, the resulting sequence is infinite). It must be noted, that recurrent sequences are not generally transformed into recurrent sequences.

Theorem 3.19. *If $x \in \mathcal{UR}$, T is a finite-state deterministic transducer and $T(x)$ is an infinite sequence, then $T(x) \in \mathcal{UR}$.*

Before we can prove our theorem, we must first note two known results.

Theorem 3.20. *If $x \in \mathcal{UR}$, $h : \Sigma^* \to \Delta^*$ is a morphism and $h(x)$ is an infinite sequence, then $h(x) \in \mathcal{UR}$.*

This theorem is a slightly adapted form of theorem 10.8.6, given in (Allouche and Shallit, 2003).

Proof. Since x is ultimately recurrent, by Definition 1.5 it can be written as $x = uz$, where z is recurrent. But then $h(x) = h(uz) = h(u)h(z)$, by Definition 1.14. Since u is finite, $h(u)$ must also be finite. To show that $h(x)$ is ultimately recurrent it is then sufficient to show that $h(z)$ is recurrent.

Suppose ν is a subsequence of $h(z)$. Then there must exist a finite subsequence v of z, that covers ν. That is, that ν is also a subsequence of $h(v)$. But, since z is recurrent, v enters z infinitely many times. Which means, $h(v)$ enters $h(z)$ infinitely many times as well. Therefore, the subsequence ν is recurrent in $h(z)$. Of course, this is true for every subsequence of $h(z)$, and therefore $h(z)$ is recurrent, which was to be shown. □

Note, that every morphism can be expressed as a transducer, but not vice versa. Since morphisms can be expressed in terms of their effects on single elements of the alphabet, it can easily be seen, that any morphism is equivalent to a transducer with a single state ($Q = \{q_0\}$).

It is obvious from definition 1.16, given a Mealy machine M, that if u is a finite sequence, then $|M(u)| = |u|$, whereas if x is an infinite sequence, then $M(x)$ also is an infinite sequence.

Theorem 3.21. *If $x \in \mathcal{UR}$ and M is a Mealy machine, then $M(x) \in \mathcal{UR}$.*

A proof of this theorem is given in (Buls, 2003).

Now we are ready to give a proof of Theorem 3.19.

Proof. (Theorem 3.19)

Since $x \in \mathcal{UR}$, it is an infinite sequence from some finite alphabet Σ. Also, there is a finite set of states Q associated with the given transducer T.

We start out by constructing a set $\Gamma = (Q \times \Sigma)^{\{0\}}$, which also must be finite, since both Q and Σ are finite. Note, that by our definition of a sequence, the elements of Γ are sequences of length 1 from the alphabet $Q \times \Sigma$. Now we can construct a Mealy

machine M from the transducer T, by changing its output alphabet to $Q \times \Sigma$ and its output function to $\eta_M = Q \times \Sigma \to \Gamma$:

$$\eta_M(q, \sigma) = (q, \sigma)^{\{0\}}, \quad q \in Q, \quad \sigma \in \Sigma \tag{3.26}$$

By Theorem 3.21, $M(x) \in \mathcal{UR}$.

Next, our goal is to construct a morphism $h : (Q \times \Sigma)^* \to \Delta^*$, such that $h(M(x)) = T(x)$, where Δ is the output alphabet of the transducer T. To do so, we define $h_1 : \Gamma \to \Delta^*$ as

$$h_1(\gamma) = \eta(\gamma(0)), \quad \gamma \in \Gamma, \tag{3.27}$$

where $\eta : Q \times \Sigma \to \Delta^*$ is the output function of T. Now we define h in terms of h_1 as

$$h(u) = \begin{cases} \lambda & : u = \lambda \\ h_1(u) & : |u| = 1 \\ h(u[0])h(u[1..(|u| - 1)]) & : |u| > 1 \end{cases} \tag{3.28}$$

It is obvious, that a h defined this way is a morphism.

From Definition 1.16 the output of T on the input x, can be expressed as the concatenation of $\eta_i = \eta(q_i, x(i))$ where $i \in \mathbb{N}$ and $q_i \in Q$ is the i-th member of the sequence of states as per Definition 1.16. On the other hand, the i-th symbol in the sequence $M(x)$ can be expressed as $\eta_M(q_i, x(i))$, since M shares its set of states with T. By the definition of h_1, however,

$$\eta(q_i, x(i)) = h_1(\eta_M(q_i, x(i)), \tag{3.29}$$

and, since h is a morphism, $T(x) = h(M(x))$.

By the formulation of the theorem, the sequence $T(x) = h(M(x))$ is infinite and, as shown previously, $M(x) \in \mathcal{UR}$ then by Theorem 3.20, $h(M(x)) = T(x)$ must be in $\mathcal{UR}$. Which was to be proved. □

4 Modularity in the semilattice of ω-words

In this chapter we show that the semilattice formed by classes of ω-words that are equivalent under transformations by Mealy machines is not modular and, therefore, not distributive.

4.1 Preliminaries

Let $(x, y) \in A^\omega \times B^\omega$. If for some Mealy machine V: $\forall n\ y[0, n] = q_0 * x[0, n]$, we say that *$V$ transforms x to y* and write $y = q_0 * x$ or $x \xrightarrow{V} y$. We write $x \rightarrow y$ if there exists such V that $x \xrightarrow{V} y$; otherwise we write $x \not\rightarrow y$. We write $x \rightleftharpoons y$ if $x \rightarrow y$ and $y \rightarrow x$ and say that *x and y are machine equivalent*; otherwise, i.e., $x \not\rightarrow y$ and $y \not\rightarrow x$, we write $x \not\rightleftharpoons y$.

4.2 Machine transformations of power-characteristic ω-words

Definition 4.1. *We will call the ω-word ${}^\zeta x \in \{0, 1\}^\omega$ the characteristic word of the power ζ if*

$${}^\zeta x(n) = \begin{cases} 1, & \text{if } \exists k \in \mathbb{N}\ n = k^\zeta; \\ 0, & \text{otherwise.} \end{cases}$$

For example, ${}^2x = 110010000100\ldots$ is the characteristic word of the squares.

Convention. Henceforth, we assume that $\zeta \geq 2$ and it is a natural number.

More generally, let $f : \mathbb{N} \to \mathbb{N}$ be any total increasing function then

$$ {}^{f}x(n) = \begin{cases} 1, & \text{if } \exists k \in \mathbb{N}\ n = f(k); \\ 0, & \text{otherwise.} \end{cases} $$

Let $V = \langle Q, \{0,1\}, \{0,1\}; \circ, * \rangle$ be a Mealy machine, where

$$ Q = \{q'_1, q'_2, \ldots, q'_b\}. $$

Applying the pigeonhole principle we can state that for every $q \in Q$ there is a least integer $i \geq 0$ such that $q \circ 0^i = q \circ 0^j$ for some $i < j$. The integer i is called the *index* of q, and $j - i$ is called the *period* of q. We can visualize this as the diagrams in the next page.

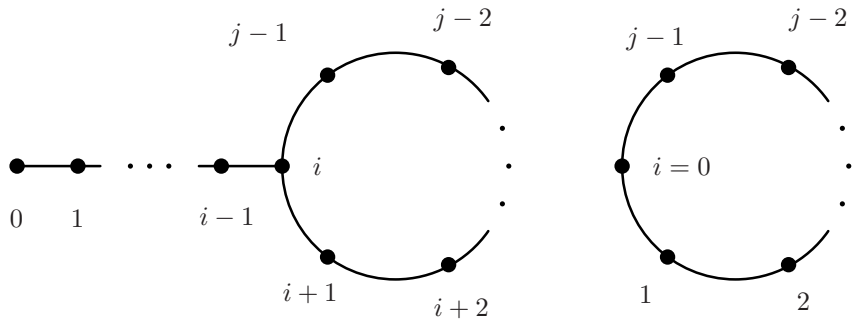

Figure 4.1: Indices and periods.

Claim 4.2. *If a_s is the index and c_s is the period of q'_s,*

$$ a_s \leq m_1 < m_2 \qquad \text{and} \qquad m_1 \equiv m_2 (\text{mod } c_s), $$

then

$$ q'_s \circ 0^{m_1} = q'_s \circ 0^{m_2}. $$

Claim 4.3. *If* $\max(a_1, a_2, \ldots, a_b) \leq m_1 < m_2$ *and* $m_1 \equiv m_2 (\mathrm{mod}\ m)$*, where*

$$m = \mathrm{lcm}(c_1, c_2, \ldots, c_b),$$

then

$$\forall q \in Q \quad q \circ 10^{m_1} = q \circ 10^{m_2}.$$

Let $\alpha(X)$ be an integer polynomial, i.e., $\alpha(X) \in \mathbb{Z}[X]$. The following theorem is known from elementary number theory.

Theorem 4.4. *If* $i \equiv j \pmod{m}$ *then* $\alpha(i) \equiv \alpha(j) \pmod{m}$.

If we take $\alpha_k = (k+1)^\zeta - k^\zeta - 1$ then we can express

$${}^\zeta x = 110^{\alpha_1}10^{\alpha_2} \ldots 10^{\alpha_k} \ldots = u_0 u_1 \ldots u_k \ldots$$

where $u_k = 10^{\alpha_k}$. Hence ${}^\zeta x[0, k^\zeta) = u_0 u_1 \ldots u_{k-1}$.

Corollary 4.5. *If* $i \equiv j \pmod{m}$ *then* $\alpha_i \equiv \alpha_j \pmod{m}$.

Let

$$\begin{aligned} w_0 &= u_0, \\ w_{k+1} &= w_k u_{k+1}. \end{aligned}$$

Then

$${}^\zeta x[0, k^\zeta) = u_0 u_1 \ldots u_{k-1} = w_{k-1}. \tag{4.1}$$

Let $q \in Q$. We define a sequence

$$q_0, q_1, \ldots, q_k, \ldots \tag{4.2}$$

where $q_k = q \circ w_k$.

Corollary 4.6. *The sequence* $q_0, q_1, \ldots, q_k, \ldots$ *is ultimately periodic.*

Proof. Let $m = \mathrm{lcm}(c_1, c_2, \ldots, c_b)$. There exists n such that

$$|u_{mn}| > \max(a_1, a_2, \ldots, a_b).$$

Now consider the sequence $q_{mn}, q_{m(n+1)}, \ldots, q_{m(n+b)}$. Since $|\overline{0, b}| = b + 1 > |Q|$ then — by the pigeonhole principle — there must exist two equal states

$$q_{m(n+i)} = q_{m(n+j)} \text{ with } 0 \leq i < j \leq b.$$

Hence

$$\begin{aligned} q_{m(n+i)+1} &= q_{m(n+i)} \circ u_{m(n+i)+1} = q_{m(n+j)} \circ u_{m(n+i)+1} \\ &\underset{\text{Claim4.3}}{=} q_{m(n+j)} \circ u_{m(n+j)+1} = q_{m(n+j)+1}. \end{aligned}$$

The rest follows by induction. □

Lemma 4.7. *Let $V = \langle Q, A, B; q_0\rangle$ be a Mealy machine. If $|Q| = m$ and $0^s \xrightarrow{V} w$ then*

$$w = uv^{\varkappa}\dot{v}, \quad \text{where} \quad |u| + |v| \le m \quad \text{and} \quad \dot{v} \in \mathrm{Pref}(v).$$

Proof. (i) If $s \le m$ then $|w| = |0^s| = s \le m$, and we can choose $u = w, v = \dot{v} = \lambda$.

(ii) Let $s > m$ and $q_0, q_1, \ldots, q_m$ are states, where

$$\forall i \in \overline{0,m} \quad q_i = q_0 \circ 0^i.$$

Since $|\overline{0,m}| = m + 1 > |Q|$ then — by the pigeonhole principle — there must exist two equal states, namely, there exist i and j, $0 \le i < j \le m$, such that

$$q_i = q_j.$$

Since $u = q_0 * 0^i$ and $v = q_i * 0^{j-i}$ it follows that

$$|u| + |v| = |0^i| + |0^{j-i}| = i + (j - i) = j \le m.$$

Choose

$$\varkappa = \left\lfloor \frac{s-i}{j-i} \right\rfloor \quad \text{then} \quad \dot{v} = q_i * 0^{s-i-\varkappa(j-i)}.$$

□

Proposition 4.8. *If ${}^{f}x \rightharpoonup y$, ${}^{\zeta}x \rightharpoonup y$ and*

$$\forall \varkappa\, \exists ak \;\; f(k) \le a^{\zeta} < (a + \varkappa)^{\zeta} \le f(k+1) \tag{4.3}$$

then y is ultimately periodic.

Proof. Since ${}^{\zeta}x \rightharpoonup y$ and ${}^{f}x \rightharpoonup y$ there exist Mealy machines

$$V = \langle Q, \{0,1\}, B; q, \circ, *\rangle \quad \text{and} \quad V' = \langle Q', \{0,1\}, B; q', \acute{\circ}, \acute{*}\rangle$$

such that ${}^{\zeta}x \xrightarrow{V} y$ and ${}^{f}x \xrightarrow{V'} y$.

(i) First, we express ${}^{\zeta}x = u_0u_1\dots u_n\dots$ with

$$u_n = 10^{\alpha_n} \quad \text{and} \quad \alpha_n = (n+1)^{\zeta} - n^{\zeta} - 1$$

and look at the sequence $q_0, q_1, \dots, q_n, \dots$ where

$$q_n = q \circ (u_0u_1\dots u_n).$$

We have shown (Corollary 4.6) that the sequence $q_0, q_1, \dots, q_n, \dots$ is ultimately periodic. Assume its period is T and the anti-period p.

(ii) By assumption (see (4.3)) we can choose integers k and a such that

$$f(k) \le a^{\zeta} < (a+T+1)^{\zeta} \le f(k+1)$$

and, moreover, $a > p+7$ and $(a+1)^{\zeta} - a^{\zeta} > 3 \cdot \max(|Q|, |Q'|) + 7$.

Now, ${}^{f}x(f(k), f(k+1))$ is a word of the form 0^d, and thus (by Lemma 4.7) $y(f(k), f(k+1))$ must be ultimately periodic with both its period and anti-period not greater than $|Q'|$. We denote this anti-period by p' and the least period by T'.

Since $p' \le |Q'| < (a+1)^{\zeta} - a^{\zeta} - 7$ then $y[(a+1)^{\zeta}, (a+1+T)^{\zeta})$ is periodic with the period T'. Notice that

$$y[(a+i)^{\zeta}, (a+i+1)^{\zeta}) = q_{a+i-1} * {}^{\zeta}x[(a+i)^{\zeta}, (a+i+1)^{\zeta}) = q_{a+i-1} * 10^{\alpha_{a+i}}$$

and that the sequence of states $q_a, q_{a+1}, q_{a+2}, \dots, q_{a+T}, \dots$ is also periodic. Therefore

$$\begin{aligned} &y[(a+i+T)^{\zeta}, (a+i+1+T)^{\zeta}) \\ &= q_{a+i-1+T} * {}^{\zeta}x[(a+i+T)^{\zeta}, (a+i+1+T)^{\zeta}) \\ &= q_{a+i-1+T} * 10^{\alpha_{a+i+T}} = q_{a+i-1} * 10^{\alpha_{a+i+T}} = \dot{u}\dot{v}\dot{w}, \end{aligned}$$

where $|\dot{u}| = |\dot{v}| = \max(|Q|, |Q'|)$. So we have two periodic words $\dot{u}\dot{v}$ and $\dot{v}\dot{w}$. Hence by Corollary 1.4

$$y[(a+i+T)^{\zeta}, (a+i+1+T)^{\zeta})$$

is periodic with period T'. Therefore we can conclude $y[(a+i)^{\zeta}, (a+i+1)^{\zeta})$ is periodic with period T' for all $i > 0$, besides, T' is the least period for all $i > 0$.

(iii) Let $X > a$ and $\mu = \operatorname{lcm}(T, T')$. Then

$$\begin{aligned} (X+\mu)^{\zeta} - X^{\zeta} &= \sum_{j=0}^{\zeta} \binom{\zeta}{j} \mu^j X^{\zeta-j} - X^{\zeta} \\ &= \mu \sum_{j=1}^{\zeta} \binom{\zeta}{j} \mu^{j-1} X^{\zeta-j} = \mu P(X), \end{aligned} \tag{4.4}$$

where

$$P(X) = \sum_{j=1}^{\zeta} \binom{\zeta}{j} \mu^{j-1} X^{\zeta-j}.$$

We have shown in (ii) that $y[X^\zeta, (X+1)^\zeta)$ is periodic. Therefore there is a v such that

$$y[X^\zeta, (X+1)^\zeta) = v^r v',$$

where $|v| = T'$ and $v' \in \mathrm{Pref}(v)$. Since T divides μ then $q_{X-1} = q_{X-1+\mu}$. But then

$$\begin{aligned} y[(X+\mu)^\zeta, (X+\mu+1)^\zeta) &= q_{X+\mu-1} *^\zeta x[(X+\mu)^\zeta, (X+\mu+1)^\zeta) \\ &= q_{X-1} *^\zeta x[(X+\mu)^\zeta, (X+\mu+1)^\zeta) \\ &= q_{X-1} * 10^{\alpha_{X+\mu}} = v^{r'} v'', \end{aligned}$$

for some number r' and $v'' \in \mathrm{Pref}(v)$. It follows from (4.4) that

$$(X+\mu+1)^\zeta - (X+\mu)^\zeta \equiv (X+1)^\zeta - X^\zeta \pmod{T'}$$

and therefore $v' = v''$.

(iv) Finally, we can select integers $\breve{k}, \breve{a}$ such that $k < \breve{k}$ and

$$f(\breve{k}) \le \breve{a}^\zeta < (\breve{a}+\mu+1)^\zeta] \le f(\breve{k}+1).$$

Now we repeat the proof from (ii). So we can conclude there is the least period $T'' \le |Q'|$ of the word $y[(\breve{a}+1)^\zeta, (\breve{a}+1+\mu)^\zeta)$. A period of

$$y[(\breve{a}+1)^\zeta, (\breve{a}+2)^\zeta)$$

is T' too. Hence (Theorem 1.3) $T'' = T'$.

Denote $y[(\breve{a}+1)^\zeta, (\breve{a}+1)^\zeta + T') = u$. Since (from formula (4.4))

$$(\breve{a}+1+\mu)^\zeta - (\breve{a}+1)^\zeta \equiv 0 \pmod{T'},$$

there is an integer s such that $y[(\breve{a}+1)^\zeta, (\breve{a}+1+\mu)^\zeta) = u^s$.

As it was shown in (iii) we can choose s_1', s_2' such that

$$y[(\breve{a}+1)^\zeta, (\breve{a}+2)^\zeta) = u^{s_1'} u' \quad \text{and} \quad y[(\breve{a}+1+\mu)^\zeta, (\breve{a}+2+\mu)^\zeta) = u^{s_2'} u'.$$

But then $y[(\breve{a}+1)^\zeta, (\breve{a}+2+\mu)^\zeta) = u^s u^{s_2'} u'$, which means that

$$y[(\breve{a}+1)^\zeta, (\breve{a}+2+\mu)^\zeta)$$

is periodic with the period T'.

Now suppose, $y[(\breve{a}+1)^{\zeta},(\breve{a}+n)^{\zeta}) = u^{\sigma}\breve{u}$, where $n > \mu+1$ and $\breve{u} \in \mathrm{Pref}(u)$. Then there exists such $\breve{v} \in \mathrm{Suff}(u)$ that $\breve{u}\breve{v} = u$. From (see formula (4.4)) $(\breve{a}+n)^{\zeta}-(\breve{a}+n-\mu)^{\zeta} \equiv 0 \pmod{T)'}$ we can conclude

$$\breve{v} \in \mathrm{Pref}\big(y[(\breve{a}+n-\mu)^{\zeta},(\breve{a}+n-\mu+1)^{\zeta})\big).$$

It follows from what we shown in (iii) that there are such σ_1, σ_2 that

$$y[(\breve{a}+n-\mu)^{\zeta},(\breve{a}+n-\mu+1)^{\zeta}) = v^{\sigma_1}v' \quad \text{and} \quad y[(\breve{a}+n)^{\zeta},(\breve{a}+n+1)^{\zeta}) = v^{\sigma_2}v',$$

with $|v| = T'$ and $v' \in \mathrm{Pref}(v)$. But then $v = \breve{v}\breve{u}$ and

$$y[(\breve{a}+1)^{\zeta},(\breve{a}+n+1)^{\zeta}) = u^{\sigma}\breve{u}v^{\sigma_2}v' = u^{\sigma}\breve{u}(\breve{v}\breve{u})^{\sigma_2}v' = u^{\sigma}(\breve{u}\breve{v})^{\sigma_2}\breve{u}v' = u^{\sigma+\sigma_2}\breve{u}v'.$$

Which means that $y[(\breve{a}+1)^{\zeta},(\breve{a}+n+1)^{\zeta})$ is periodic with period T'.

Now, by induction, we have $y[(\breve{a}+1)^{\zeta},(\breve{a}+i)^{\zeta}]$ is periodic with the period T' for any $i > 1$. Hence, y is ultimately periodic. □

4.3 Modularity in the semilattice of ω-words

Our main object of investigation is the machine poset of infinite words. In order to avoid some set-theoretical problems we make some assumptions. Let us take the set $\mathfrak{N} = \bigcup_{k=0}^{\infty} \big(\overline{0,k}\big)^{\omega}$. We shall assume that the states of the involved Mealy machines as well as their input and output alphabets all are from the set $\mathbb{N}$. If another input or output alphabet A is used, we assume that there exists a bijection $\beta : A \to \overline{0,|A|-1}$ and that this bijection is applied to the input or output word, respectively.

We suppose the reader is familiar with the basic notions of ordered sets (Birkhoff, 1967). If $\rightharpoonup$ is used as an algebraic relation on $\mathfrak{N}$, then the algebraic structure $\langle \mathfrak{N}, \rightharpoonup \rangle$ defines a preorder (Belovs, 2008), while the quotient set $\tilde{\mathfrak{N}} = \mathfrak{N}/\rightleftharpoons$ becomes the ordered set $\langle \tilde{\mathfrak{N}}, \rightharpoonup \rangle$. It has been shown that this poset $\tilde{\mathfrak{N}}$ is a join-semilattice (Belovs, 2008), where the join $[(x_i)] \vee [(y_i)] = [(x_i, y_i)]$.

Definition 4.9. *A join-semilattice $\langle D, \leq \rangle$ is distributive when*

$$\forall xab\ (\ x \leq a \vee b\ \Rightarrow\ \exists a'b'\ (\ a' \leq a\ \&\ b' \leq b\ \&\ x = a' \vee b'\ ))$$

A join-semilattice $\langle D, \leq \rangle$ is modular when

$$\forall xab\ (\ a \leq x \leq a \vee b\ \Rightarrow\ \exists b' \leq b\ (x = a \vee b'\ ))$$

Theorem 4.10. *The join-semilattice $\langle \tilde{\mathfrak{N}}, \rightharpoonup \rangle$ is not modular.*

Proof. We start by showing that ${}^2x \vee {}^4x \rightharpoonup x'$, where

$$x'(n) = \begin{cases} 1, & \text{if } \exists k \in \mathbb{N}\ n = k^4; \\ 1, & \text{if } \exists k \in \mathbb{N}\ n = (k^2+1)^2; \\ 0, & \text{otherwise.} \end{cases}$$

By definition $({}^2x \vee {}^4x)(n) = ({}^2x(n), {}^4x(n))$. Define the Mealy machine

$$V = \langle \{q_0, q_1, q_2\}, \{\tbinom{0}{0}, \tbinom{0}{1}, \tbinom{1}{0}, \tbinom{1}{1}\}, \{0,1\}; q_0, \circ, * \rangle$$

by

$$\begin{aligned} q_1 &= q_0 \circ \tbinom{0}{0} = q_0 \circ \tbinom{0}{1} = q_0 \circ \tbinom{1}{0} = q_0 \circ \tbinom{1}{1} = q_1 \circ \tbinom{0}{0} = q_1 \circ \tbinom{0}{1} = q_1 \circ \tbinom{1}{0} = q_2 \circ \tbinom{1}{0}, \\ q_2 &= q_1 \circ \tbinom{1}{1} = q_2 \circ \tbinom{0}{0} = q_2 \circ \tbinom{0}{1} = q_2 \circ \tbinom{1}{1}; \\ 0 &= q_1 * \tbinom{0}{0} = q_1 * \tbinom{0}{1} = q_1 * \tbinom{1}{0} = q_2 * \tbinom{0}{0} = q_2 * \tbinom{0}{1} = q_2 * \tbinom{1}{1}, \\ 1 &= q_0 * \tbinom{0}{0} = q_0 * \tbinom{0}{1} = q_0 * \tbinom{1}{0} = q_0 * \tbinom{1}{1} = q_1 * \tbinom{1}{1} = q_2 * \tbinom{1}{0}. \end{aligned}$$

We illustrate this by the diagram in Figure 4.2.

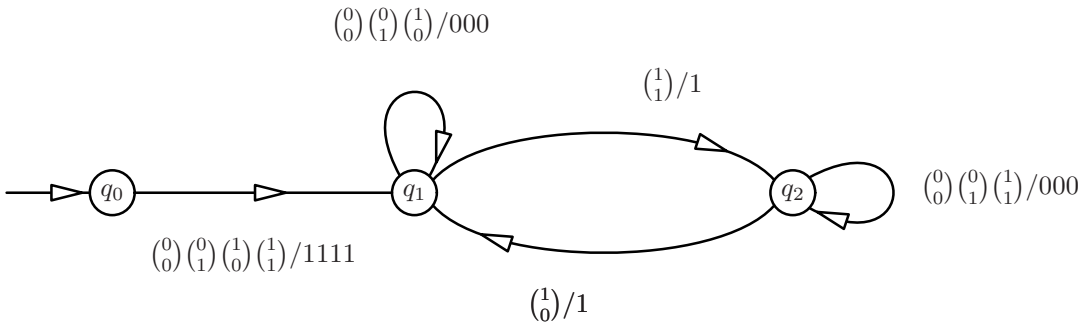

Figure 4.2: ${}^2x \vee {}^4x \xrightarrow{V} x'$

It follows straightforwardly from the construction that ${}^2x \vee {}^4x \xrightarrow{V} x' \rightharpoonup {}^4x$. Now suppose there exists y such that ${}^2x \rightharpoonup y$ and $x' \rightleftharpoons {}^4x \vee y$. But then $x' \rightharpoonup y$ too. Notice $x' = {}^g x$ for

$$g(k) = \begin{cases} 0, & \text{if } k = 0; \\ \frac{(k+1)^4}{16}, & \text{if } k \text{ is odd;} \\ (\frac{k^2}{4}+1)^2, & \text{if } k \text{ is even.} \end{cases}$$

Hence, by Proposition 4.8 y is ultimately periodic. But if so, then ${}^4x \vee y \rightleftharpoons {}^4x$. However this is a contradiction, because then $x' \rightleftharpoons {}^4x$. □

Corollary 4.11. *The join-semilattice* $\langle \tilde{\mathfrak{N}}, \rightharpoonup \rangle$ *is not distributive.*

We recall that every distributive join-semilattice is modular.

Conclusions

The main result of this thesis answers a decision problem in finitely generated bi-ideals by giving an effective procedure for determining if two finite bases generate the same bi-ideal. My work is a continuation of the work begun by (Buls and Lorencs, 2008) and (Lorencs, 2012). While this result completely resolves the problem in finitely generated bi-ideals, it remains to be seen whether our approach can be applied to other classes of bi-ideals, such as restricted bi-ideals, for example.

Some smaller results obtained in the thesis shed some light on the properties of finitely generated bi-ideals. I show that they are a subclass of morphic words and that they are closed under transformation by morphism and left shift. We also investigated some properties of bi-ideals in the context of possible cryptographic uses. We find that the S-sequence of a shrinking generator can be replaced by a finitely generated bi-ideal, without breaking its desirable statistical properties. We give some conditions under which the resulting pseudo-random sequence is aperiodic, including showing a subclass of bi-ideals for which this holds for any non-trivial A-sequence. This, however, can only be considered the very beginning of a research programme, since no general condition for the aperiodicity of shrunk sequences is found.

Finally, we also shortly consider the semilattice of machine-invariant language classes as introduced by (Buls, 2003). We demonstrate that this semilattice is not modular (and, therefore, not distributive).

Bibliography

Allouche, J.P. and J. Shallit. 1999. *The ubiquitous prouhet-thue-morse sequence.*

_____. 2003. *Automatic sequences: theory, applications, generalizations*, Cambridge Univ Pr.

Belovs, A. 2008. *Some Algebraic Properties of Machine Poset of Infinite Words*, RAIRO-Theoretical Informatics and Applications **42**, no. 3, 451–466.

Berzina, I., R. Bets, J. Buls, E. Cers, and L. Kulesa. 2011. *On a non-periodic shrinking generator*, proceedings of the 13^{th} symposium on symbolic and numeric algorithms for scientific computing (SYNASC 2011), pp. 348–354.

Birkhoff, G. 1967. *Lattice theory, vol. 25*, American Mathematical Society Colloquium Publications.

Büchi, J.R. 1960. *Weak second-order arithmetic and finite automata*, Mathematical Logic Quarterly **6**, no. 1-6, 66–92.

Buls, J. 2003. *Machine Invariant Classes*, Proceedings of WORDS **3**, 10–13.

_____. 2005. *The Lattice of Machine Invariant Sets and Subword Complexity*, Arxiv preprint cs/0502064.

Buls, J. and A. Lorencs. 2008. *From bi-ideals to periodicity*, RAIRO-Theor. Inf. Appl. **42**, no. 3, 467–475.

Cers, E. 2008. *The properties of bi-ideals in the frequency test*, Tatra Mt. Math. Publ **41**, 107–117.

Cobham, A. 1972. *Uniform tag sequences*, Theory of Computing Systems **6**, no. 1, 164–192.

Coppersmith, D., H. Krawczyk, and Y. Mansour. 1994. *The shrinking generator*, Advances in cryptology – crypto93, 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.

Dekking, F.M. 1994. *Iteration of maps by an automaton*, Discrete Mathematics **126**, no. 1-3, 86.

Diekert, V. and M. Kufleitner. 2011. *Fragments of first-order logic over infinite words*, Theory of Computing Systems **48**, 486–516.

Fine, N.J. and H.S. Wilf. 1965. *Uniqueness theorem for periodic functions*, Proc. Amer. Math. Soc. **16**, 109–114.

Holton, C. and L.Q. Zamboni. 2000. *Iteration of maps by primitive substitutive sequences*, 137.

Kleene, S.C. and E.L. Post. 1954. *The upper semi-lattice of degrees of recursive unsolvability*, The Annals of Mathematics **59**, no. 3, 379–407.

Lorencs, A. 2012. *The identity problem of finitely generated bi-ideals*, Acta Informatica, 1–11.

Lyndon, R.C. and M.P. Schützenberger. 1962. *The equation $a^M = b^N c^P$ in a free group*, Michigan Math. J **9**, no. 4, 289–298.

Marsaglia, G. 1996. *Diehard: a battery of tests of randomness*, See http://stat. fsu. edu/ geo/diehard. html.

Mealy, G.H. 1955. *A method for synthesizing sequential circuits*, Bell System Technical Journal **34**, no. 5, 1045–1079.

Morse, H.M. 1921. *A one-to-one representation of geodesics on a surface of negative curvature*, American Journal of Mathematics **43**, no. 1, 33–51.

Morse, M. and G.A. Hedlund. 1940. *Symbolic dynamics ii. sturmian trajectories*, American Journal of Mathematics **62**, no. 1, 1–42.

Muchnik, A., A. Semenov, and M. Ushakov. 2003. *Almost periodic sequences*, Theoretical Computer Science **304**, no. 1-3, 1–33.

Neuenschwander, D. 2004. *Probabilistic and statistical methods in cryptology: an introduction by selected topics*, Springer-Verlag New York Inc.

Oishi, S. and H. Inoue. 1982. *Pseudo-random number generators and chaos*, IEICE transactions **1976**.

Patidar, V. and K.K. Sud. 2009. *A novel pseudo random bit generator based on chaotic standard map and its testing*, EJTP **6**, no. 20, 327–344.

Perrin, D. and J.E. Pin. 2002. *Infinite words*, Elsevier/Academic Press.

Phatak, S.C. and S.S. Rao. 1995. *Logistic map: A possible random-number generator*, Physical review E **51**, no. 4, 3670.

Restivo, A. and C. Reutenauer. 1984. *On the burnside problem for semigroups*, J. Algebra **89**, no. 1, 102–104.

Sandri, G.H. 1992. *A simple nonperiodic random number generator: A recursive model for the logistic map*, DTIC Document.

Schmidt, W.M. 1980. *Diophantine approximation*, Springer Verlag.

Schneier, B. 1995. *Applied cryptography: Protocols*, Algorithms, and Source code in C, 574–577.

Simon, I. 1988. *Infinite words and a theorem of hindman*, Rev. Mat. Apl **9**, 97–104.

Zimin, A.I. 1982. *Blocking sets of terms*, Matematicheskii Sbornik **161**, no. 3, 363–375.

Author's publications

Bērziņa I., R. Bēts, J. Buls, E. Cers and L. Kuleša. 2011. *On a non-periodic shrinking generator,* proceedings of the 13^{th} International symposium on symbolic and numeric algorithms for scientific computing (SYNASC 2011), IEEE Computer Society, 348–354.

Buls J. and E. Cers. 2010. *Distributivity in the semilattice of ω-words,* Contributions to General algebra **19**, 13–22.

Buls J. and E. Cers. 2010. *Modularity in the semilattice of ω-words,* proceedings of the 13^{th} Mons theoretical computer science days (JM 2010), Université de Picardie Jules Verne.

Cers, E. 2008. *The properties of bi-ideals in the frequency test,* Tatra Mt. Math. Publ **41**, 107–117.

Cers, E. 2010. *An unique basis representation of finitely generated bi-ideals,* proceedings of the 13^{th} Mons theoretical computer science days (JM 2010), Université de Picardie Jules Verne.