# A GRADUATE SEMINAR REPORT ON REPETITION IN WORDS

By
Yassin Tesfaw

አዲስ አበባ ዩኒቨርሲቲ
Addis Ababa University

In partial fulfillment of the requirment for the
Degree of Masters of Science in Mathematics

DEPARTMENT OF MATHEMATICS

Advisor: Yirgalem Tsegaye(PhD)

July 2, 2013

ADDIS ABABA UNIVERSITY
DEPARTMENT OF MATHEMATICS

The undersigned hereby certify that they have read and recommend to the School of Graduate Studies for acceptance a project work entitled "**REPETITION IN WORDS**" by **YASSIN TESFAW** in partial fulfillment of the requirements for the degree of Master of Science.

date____________________________

Adivisor Signature

Dr.Yirgalem Tsegaye ___________________

Examiners Signature

1. ____________________________ ___________________

2. ____________________________ ___________________

# Contents

# Summary

Combinatorics on words has imposed itself as a powerful tool for the study of large number of discrete, linear, non-commutative objects. This report project is intended to discuss words, in particular, repetition in words and avoiding repetition of squares, overlaps and cubes over binary and ternary alphabet of infinite words and Abelian repetition.

# Acknowledgement

I would like to express my gratitude to my advisor Dr.Yirgalem Tsegaye for her valuable guidance and support. I would particularly like to thank her for introducing me to the area of combinatorics on words. Thanks are also due to the Department of Mathematics for material support.

# Introduction

Words are the central notion of this project report focuses mainly repetition in words. A word is a finite or infinite sequence of symbols taken from a finite alphabet.

In Chapter 1, we give notations and basic definitions of word, morphism and power of words. In Chapter 2 we give an overview of repetition in words specially on avoiding repetition in infinite word and abelian repetition and thier avoidability in infinite word by using Dekking[7] construction.

The study of combinatorics on words dates back at least to the beginning of the $20^{th}$ century and the work of Axel Thue[6] in 1906-12. Thue's work on combinatorics on words was largely concerned with repetitions in words, and his primary technique for studying such repetitions was the use of iterated morphisms. When A.Thue started to work on repetition-free words. He proved, among other things, the existence of an infinite square-free word over a ternary alphabet.

Once the foundation of the theory was laid down it developed rapidly. In 1983, Lothaire wrote the first book on combinatorics on words in 1983. This book had an enormous influence on the further development of the field.

# Chapter 1

# Preliminaries

In this chapter, we introduce the basic terminologies and results that we need in subsequent chapters. We will begin with defining semigroup the most basic concepts, like letters and words(finite or infinite(one sided right)), and then proceed to the more advanced notions of morphic sequences, morphism, fixed point of morphism and powers of words.

## 1.1 Semigroup

A **semigroup** is a nonempty set equipped with binary associative operation. Let $S$ be a semigroup and $\emptyset \neq T \subset S$. Then $T$ is a subsemigroup of $S$ if $a, b \in T \Rightarrow ab \in T$.
Let $S \ \ and \ \ T$ be semigroups, then $f : S \to T$ is a semigroup morphism if, $f(uv) = f(u)f(v)$ for all $u, v \in S$.
A **Monoid** $M$ is a semigroup with an identity element. i.e an element $e$ such that $me = em = m$ for all $m \in M$.
If $M$ is monoid then $N$ is a submonoid of $M$ if $N$ is a subsemigroup and $e \in N$. If $M$ and $N$ are monoids then $f$ is a monoid morphism if $f$ is a semigroup morphism and $f(e_M) = e_N$.
Given a semigroup $S$ and a set $X \subset S$, that is $X^+ = \{x_1 x_2 \cdots x_n | n \geq 1, x_i \in X\}$ the subsemigroup generated by $X$. If $S$ is a monoid we also define $X^* = X^+ \cup \{e\}$ which is a submonoid generated by $X$.
A monoid $M$ is free if $X \subset M$ such that:

1. $M = X^*$, and
2. For all $n, m \geq 0$ and $x_1, x_2, \cdots, x_n, y_1, y_2, \cdots, y_m \in X$, we have

$$x_1 \cdots x_n = y_1 \cdots y_m \ \ implies \ \ n = m \ \ and \ \ x_i = y_i \ \ for \ \ i = 1, 2, \cdots n$$

Condition (1) means that $X$ generates $M$, i.e is a generating set of $M$, and condition (2) requires that each element of $M$ has the unique representation as a product of element of $X$. The subset $X$ of $M$ satisfying (1) and (2) is called a base of $M$.

## 1.2 Word

An **Alphabet** is finite set of symbols, the element of $A$ is called a letter.
A **Word** is a sequence of symbols from the alphabet. Words can be finite or infinite (to the right).
Empty word is a sequence of zero symbols. An empty word is denoted by $\epsilon$.
The set of all finte words over alphabet $A$ is denoted by $A^*$.
The set of all finite nonempty words over alphabet $A$ is denoted by $A^+$, then $A^+ = A^* \backslash \{\epsilon\}$.
**Catenation** or product of words is an operation defined on $A^*$ by

$$a_la_2\cdots a_n \cdot b_lb_2\cdots b_n = a_la_2\cdots a_nb_lb_2\cdots b_n,\ for\ \ a_i, b_i \in A$$

Catenation of words is associative. i.e

$$(a_1\cdots a_n.b_1\cdots b_m).c_1\cdots c_t = (a_1\cdots a_nb_1\cdots b_m).c_1\cdots c_t = a_1\cdots a_nb_1\cdots b_nc_1\cdots c_n$$

And

$$a_1\cdots a_n.(b_1\cdots b_m.c_1\cdots c_t) = a_1\cdots a_n.(b_1\cdots b_mc_1\cdots c_t) = a_1\cdots a_nb_1\cdots b_mc_1\cdots c_t$$

The empty word is the identity element with respect to this operation.
Consequently $A^* = (A^*, \cdot)$ and $A^+ = (A^+, \cdot)$ are a **monoid** and a **semigroup** respectively and $A^*,\ A^+$ are free monoids and free semigroups resectively generated by $A$.

**Example 1** *For any alphabet $A$(not only finite ones) the monoid $A^*$(or the semigroup $A^+$) is free with the base $A$. These are called the free monoids and semigroups generated by $A$.*

If $w \in A^*$, then there exist a unique integer $n \geq 0$ and letters $a_l, a_2, \cdots, a_n \in A$, such that $w = a_la_2\cdots a_n$, the number $n$ is called **length** of $w$ denoted by $|w|$. i.e. the total number of letters in $w$.
If $w \in A^*$ and $a \in A$, then $|w|_a$ denotes the number of occurrence of letter $a$ in the word $w$. so that:

$$|w| = \sum_{a\in A} |w|_a$$

Alph(w) is the set of letters having at least one occurrence in the word. i.e $Alph(w) = \{a \backslash |w|_a \geq 1\}$.

**Example 2** *Let* $A = \{a, b, c\}$ *and* $W = abacab$ *where* $w \in A^*$ *then* $|w|_a = 3, |w|_b = 2, |w|_c = 1$ *and* $|w| = 6 = 3 + 2 + 1$

Let $\mathbf{N}$ denote the set $\{0, 1, 2, ...\}$. An **infinite (one sided right) word** is a map from $\mathbf{N}$ to $A$, typically written as a sequence such as

$$w = w(0)w(1)\cdots \quad or \quad w = w_0w_1w_2\cdots$$

The set of all infinite word is denoted by $A^\omega$. $A^\infty = A^* \cup A^\omega$ to denote the set of finite or infinite words. If $w$ is finite nonempty word, then $w^\omega$ denotes the infinite word $www\cdots$.
An infinite word $w$ is **ultimately periodic** if it can be written in the form $xy^\omega$ with $y$ non empty and $x, y \in A^*$.
A word $w'$ is called a **factor(resp.left factor or prefix, a right factor or suffix** ) of a word $w$ if there exist words $x, y$ such that $w = xw'y(resp.\ \ w = w'y, w = xw')$. The factor (resp. prefix, suffix) is **proper** if $xy \neq \epsilon$ (resp. $y \neq \epsilon, x \neq \epsilon$).
The set of all prefixes of $w$ is denoted by $pref(w)$, while $pref_k(w)$ means the prefix of length $k$ of $w$ (or $w$ if $|w| < k$). Similarly, by $suf(w)$, we mean the set of suffixes of $w$.
Let $w \in A^+$ and $w = a_1a_2\cdots a_n$ with $a_i \in A$. A period of $w$ is an integer $P$ such that

$$a_i = a_{i+p} \quad for \quad i = 1, 2, \cdots, n - p \tag{1.1}$$

The smallest $p$ of $w$ satisfying (1.1) is called period of $w$ and denoted by $p(w)$.
Note that: A word have several periods and any $q \geq n$ is a period of $w$.

**Example 3** *Let* $A = \{a, b\}$ *then words* ***abababa*** *and* ***aabaabbaabaa*** *have periods* $2, 4, 6 \quad and \quad 7, 10, 11,$ *respectively.*

A **factorization** of a word $w$ is any sequence $u_1, u_2, \cdots, u_n$ of words satisfying

$$w = u_1u_2\cdots u_n \tag{1.2}$$

$w$ is $X$-factorization if all $u_i's$ are from $X$. It is natural to write:

$$X^* = \{u_1u_2\cdots u_n | n \geq 0 \quad and \quad u_i \in X\}$$

$$X^+ = \{u_1u_2\cdots u_n | n \geq 1 \quad and \quad u_i \in X\}$$

The sets $X^*$ and $X^+$ are submonoids and subsemigroups of $A^*$ and $A^+$ respectively, so-called submonoids and subsemigroups generated by $X$.

**Example 4** *Let* $X = \{a, ab, \}$, *then* $X^*$ *is a monoid, a submonoid of* $A^*$ *and a submonoid generated by* $X$.

The **reversal** of a word $w = a_l a_2 \cdots a_n$, $a_i \in A$, is the word $\tilde{w} = a_n \cdots a_2 a_1$

A word $w$ is **palindrome** if $w = \tilde{w}$.
If $|w|$ is even then, $w$ is a palindrom if and only if $w = x\tilde{x}$ fore some word $x$.

**Proof 1** ($\Rightarrow$) *Assume that* $w$ *is a palindrome with even length. We want to show that* $w = x\tilde{x}$. *Let* $w = a_0 a_1 a_2 a_3 \ldots a_n$, *and* $a_i \in A$ *with* $n$ *even. Then* $\tilde{w} = a_n a_{n-1} a_{n-2} a_{n-3} \ldots a_0$. *Since* $w$ *is a palindrome we have* $w = a_0 a_1 a_2 a_3 \ldots a_n = a_n a_{n-1} a_{n-2} a_{n-3} \ldots a_0 = \tilde{w}$. *From this* $a_0 = a_n$, $a_1 = a_{n-1}, \cdots, a_k = a_{n-k}, a_{k+1} = a_{n-(k+1)}, \cdots, a_n = a_{n-n} = a_0$, *where* $k = \frac{n}{2}$. *Then* $|x| = \frac{n}{2} = k$ *and* $x = a_0 a_1 \cdots a_k = a_k a_{k+1} \cdots a_n = \tilde{x}$. *hence* $w = x\tilde{x}$.
($\Rightarrow$) *Assume* $w = x\tilde{x}$, *then we want to show that* $w$ *is a palindrome i.e.* $w = \tilde{w}$.
*Then* $\tilde{w} = (x\tilde{x})\tilde{} = x\tilde{x} = w$.

Other wise $w$ is a palindrome if and only if $w = xa\tilde{x}$ for some word $x$ and a letter $a$. For example the English word *rotator* and *deed* are palindromes. For any word $w$ over the binary alphabet $\{a, b\}$, we denote by $\overline{w}$ the **complement** of $w$, namely the word obtained from $w$ by changing $a'$s to $b'$s and $b'$s to $a'$s.
For any word $w \in A^*$, the square of a word $w$ is a monoid product $w^2 = ww$. To pass from the finite to the infinite, we shall often rely (implicitly) on the following form of a result of Konig [4] known as the Infinity Lemma.

**Theorem 1.2.1** *(Konigs Infinity Lemma). Let* $A$ *be a finite alphabet. Let* $B$ *be an infinite subset of* $A^*$. *There exists an infinite word* $w$ *such that every prefix of* $w$ *is a prefix of at least one word in* $B$.

**Proof 2** *There must exist a letter* $w_o \in A$, *such that infinitely many words in* $B$ *begin with* $w_o$. *Similarly there must exist a letter* $w_1 \in A$, *such that infinitely many words in* $B$ *begin with* $w_o w_1$. *Counting in this fashion one defines an infinite word* $w = w_o w_1 w_2 w_3 \cdots$ *such that every prefix of* $w$ *is a prefix of at least one word in* $B$.

## 1.3 Morphic Sequence

In this section we first define one of the most fundamental tools in combinatorics on words, a morphism. Then we define morphic sequences and fixed points of morphism and give some examples. Finally, we define the incidence matrix of a morphism.

### 1.3.1 Morphism

The notion of a morphism is fundamental to combinatorics on words in general.

**Definition 1.3.1** *A map* $h : A^* \to B^*$ *where* $A, \quad B$ *are finite alphabets is called a morphism if* $h$ *satisfies* $h(uv) = h(u)h(v)$ *for all* $u, v \in A^*$.

In particular it follows that

- $h(\epsilon) = \epsilon$
- A morphism may be specified by providing the image word $h(a)$ for all $a \in A$.

**Example 5** *Define a morphism* $h : \{a, b, c\}^* \to \{a, b, c\}^*$ *by:*

$$a \to abcab, \quad b \to acaabc, \quad c \to acbcacb$$

*take* $u = cb, v = a$ *where* $u, v \in A^*$, *then* $h(uv) = h(u)h(v) = h(cb)h(a) = h(c)h(b)h(a) = acbcacbacaabcabcab$

Below is a list for some types of morphisms.

A morphism $h : A^* \to A^*$ is:

1. **Prolongable** on a letter $a \in A$, if $h(a) = ax$ for some word $x \in A^+$ such that $h^n(x) \neq \epsilon$ for all integers $n \geq 1$.

2. **K-uniform**, where $k \geq 2$ is an integer, if $|h(a)| = k$ for all letters $a \in A$, it is uniform if it is k-uniform for some $k$.

3. **Non-erasing** if $h(a) \neq \epsilon$ for all letters $a \in A$.

**Example 6** *In Example 5, the morphism $h$ is non-erasing and prolongable on a letter $a$ since $h(a) = abcab$ where $bcab \in A^+$ and $h(a) \neq \epsilon$ for all letters $a \in A$. But it is not $K$-uniform because $|h(a)| \neq |h(b)|$.*

**Definition 1.3.2** *Let $h : A^* \to A^*$ be a morphism. A finite or infinite word $w$ such that $h(w) = w$ is said to be a fixed point of $h$.*

**Definition 1.3.3** *A sequence $(u_n)_{n\geq 0}$ of finite words over an alphabet $A$ converges to an infinite word $X$, if every prefix of $X$ is a prefix of all but finite number of terms. This word is unique and is denoted by*

$$\lim_{n\to\infty} u_n = X$$

**Example 7** *The sequence $a^n b^n$ converges to the infinite word $a^\omega$.*
*Given the sequence $(u_n)_{n\geq 1} = a^n b^n$ and $X = x_0 x_1 \cdots x_n \cdots$ be an infinite word where $x_n$ is a letter for all $a \in A$. Our claim is to show $a^n b^n$ converges to the infinite word $a^\omega$. First let us denote the prefix of length $k \geq 0$ of $X$ is $X^{[k]} = x_0 x_1 \cdots x_{k-1}$ and $u_1 = ab, u_2 = aabb, \cdots, u_n = a^n b^n$. Then from the definition $X^{[k]} = u_n$ where $k = |u_n|$ for $n \geq 1$ and $X^{[k]}$ is a prefix of all $u_m$ for $m \geq n$. But $a^n$ is the only prefix of $X$ that is a prefix of all $u_m$ for $m \geq n$. hence*

$$\lim_{n\to\infty} u_n = \lim_{n\to\infty} a^n b^n = a^\omega = aaaaaa\cdots$$

An important special case arises when $u_n$ is a prefix of $u_{n+1}$ then the sequence converges provided the length of the word $u_n$ is unbounded. A special case of this describes in the following proposition.

**Proposition 1** *Let $h$ be a nonerasing morphism from $A^*$ into itself, and let $a$ be a letter such that $h(a) = as$ for some nonempty word $s$. set for $n \geq 0$,*

$$u_n = h^n(a) \quad \textit{and} \quad v_n = h^n(s)$$

*then*

1. *$u_{n+1} = u_n v_n$ and inparticular $u_n$ is a prefix of $u_{n+1}$ for all $n > 0$.*
2. *$u_{n+1} = av_0 v_1 \cdots v_n$.*
3. *The infinite word*

   $$X = ah^0(s)h^1(s)h^2(s)\cdots h^n(s)\cdots \tag{1.3}$$

   *is the limit of the word $u_n$ and $X$ is a fixed point of $h$. Moreover it is a unique fixed point of $h$ starting with letter $a$.*

**Proof 3** *1.* $u_{n+1} = h^{n+1}(a) = h^n(h(a)) = h^n(a)h^n(s) = u_n v_n$.

*2. By using induction for* $n = 0, u_1 = h^1(a) = as = av_0$, *then assume it is true for* $n$.

$$i.e \quad u_n = h^n(a) = ah^0(s)h^1(s)\cdots h^{n-1}(s) = av_0v_1\cdots v_{n-1}$$

*now we want to show it is true for* $n+1$

$$u_{n+1} = h^{n+1}(a) = h^n(h(a)) = h^n(as) = h^n(a)h^n(s) = h^n(a)v_n = av_0v_1\cdots v_{n-1}v_n = u_{n+1}$$

*3. By definition* $\lim_{n\to\infty} u_n = X = \lim_{n\to\infty} h^n(a) = ash(s)h^2(s)\cdots h^n(s)\cdots$ *and* $h(X) = h(ash(s)h^2(s)\cdots h^n(s)\cdots) = h(a)h(s)(s)h^2(s)\cdots h^n(s)\cdots = ash(s)h^2(s)\cdots h^n(s)\cdots = X$ *which is a fixed point of* $h$.
*Uniqueness:*
*Let* $Y$ *be an infinite word, since* $h(a) \neq \epsilon$ *for all* $a \in A$, $h(Y)$ *is also an infinite word.*

$$h(X) = X \quad for \quad X = \lim_{n\to\infty} h^n(a)$$

*In other term* $X$ *is a fixed point of* $h$.
*Let* $Y = h(X)$, *for each prefixes* $u$ *of* $X$, *the word* $h(u)$ *is a a prefix of* $h(X)$. *Thus each* $h^n(a)$, $n > 0$ *is a prefix of* $Y$, *and* $Y$ *starts with* $a$. *then* $Y = \lim h^n(a) = X$.

The word $X$ of the proposition is denoted by $h^\omega(a)$

**Definition 1.3.4** *The sequence* $u_n = h^n(a)$ *is called Morphic Sequence and the word* $X = h^\omega(a)$ *is called morphic word.*

**Example 8** *The infinite word* $\alpha$ *over the alphabet* $A = \{a, b\}$ *is defined as the limit*

$$\alpha = \lim_{n\to\infty} u_n$$

*Where the sequence of words* $(u_n)_{n\geq 0}$ *and* $(v_n)_{n\geq 0}$ *are defined by*

$$u_0 = a \qquad v_0 = b$$

$$u_{n+1} = u_n v_n \qquad v_{n+1} = v_n u_n, \quad n \geq 0$$

*The first letters of* $\alpha$ *are*

$$\alpha = \mu^\omega(a) = abbabaabbaababbaba\cdots$$

*The word* $\alpha$ *is a morphic word since* $u_n = \mu^n(a)$ *where* $\mu$ *is a morphism* $\mu : \{a, b\}^* \to \{a, b\}^*$ *defined by:*

$$a \to ab$$

$$b \to ba$$

*The decomposition of* $X$ *corresponding to equation* (1.3) *is :*

$$\alpha = \mu^\omega(a) = a\ b\ ba\ baab\ baababba \cdots$$

Defining infinite words in this fashion, as fixed points of morphisms, is a useful technique that will be employed often in what follows. Let us underscore the necessary ingredients. Fix $n \in \mathbb{N}$ and a morphism $h : A^* \to A^*$. We write $h^n$ for the $n$-fold composition of $h$ with itself, the $n$-th **iterate** of $h$. If $a$ is a prefix of $h(a)$ for a given $a \in A$, then $h^n(a)$ is a prefix of $h^{n+1}(a)$ for all positive integers $n$: indeed, writing $h(a) = au$, we have

$$h^{n+1}(a) = h^n(h(a)) = h^n(au) = h^n(a)h^n(u)$$

Therefore, the sequence $h^1(a), h^2(a), h^3(a), \cdots$ has a (unique) well-defined limit, which we denote by

$$\lim_{n\to\infty} h^n(a) = h^\omega(a)$$

**Example 9** *In the above example the sequence* $\mu^1(a), \mu^2(a), \mu^3(a), \cdots$ *has well defined limit, which is denoted by*

$$\mu^\omega(a) = \lim_{n\to\infty} \mu^n(a)$$

**Definition 1.3.5** *In the example above* $\alpha$ *is called the Thue-Morse word and* $\mu$ *is called the Thue-Morse Morphism.*

### 1.3.2 The Incidence Matrix of a Morphism

**Definition 1.3.6** *Let* $h : A^* \to B^*$ *be a morphism for some finite set* $A = a_1, a_2, \cdots a_d$*, Then we define incident matrix* $M = M(h)$ *associated with the morphism* $h$ *is the matrix given by:*

$$M = (m_{i,j})_{1 \leq i,j \leq d}$$

*Where* $m_{i,j}$ *is the number of occurrences of* $a_i$ *in* $h(a_j)$*, i.e.* $m_{i,j} = |h(a_j)|$

Then $M$ can be written as:

$$\mathbf{M(h)} = \begin{pmatrix} |h(a_1)|_{a_1} & |h(a_2)|_{a_1} & \cdots & |h(a_j)|_{a_1} \\ |h(a_1)|_{a_2} & |h(a_2)|_{a_2} & \cdots & |h(a_j)|_{a_2} \\ \vdots & \vdots & \cdots & \vdots \\ |h(a_1)|_{a_i} & |h(a_2)|_{a_i} & \cdots & |h(a_j)|_{a_i} \end{pmatrix}$$

**Example 10** *consider the morphism defined by* $h : \{a, b, c\}^* \to \{a, b, c\}^*$ *defined by* $a \to ab, \quad b \to cc, \quad c \to bb$*, then the incident matrix is*
$M(h)$

$$\mathbf{M(h)} = \begin{pmatrix} |h(a)|_a & |h(b)|_a & |h(c)|_a \\ |h(a)|_b & |h(b)|_b & |h(c)|_b \\ |h(a)|_c & |h(b)|_c & |h(c)|_c \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 2 \\ 0 & 2 & 0 \end{pmatrix}$$

### 1.3.3 Powers of Words

**Definition 1.3.7** *For any positive integer* $k \geq 2$*, a* $k$*- power is a nonempty word of the form* $\overbrace{uu \cdots u}^{k}$ *written convenience as* $u^k$*. A 2- power is said to be a square and 3- power is a cube.*

**Example 11** *The English word bonbon is a square word.*

**Definition 1.3.8** *A nonempty word that is not a* $k$*- power for any* $k \geq 2$ *is called primitive word. i.e.* $w = u^k$ *implies* $k = 1$ *for* $w, u \in A^+$*.*

**Definition 1.3.9** *An overlap is a word of the form* $avava$*, where* $a \in A$ *and* $v \in A^*$*.*

**Example 12** *The English word alfalfa*

Now we extend the notion of integer power in to fractional powers.

**Definition 1.3.10** *Let $k \in \Re$ and $k > 1$:*

1. *A word $w$ is said to be $k$- power of $u$, if $w$ is the shortest prefix of $u^\omega$ such that $|w| \geq k|u|$.*
2. *A word $w$ is said to be an $k^+$- power of $u$, if $w$ is the shortest prefix of $u^\omega$ such that $|w| > k|u|$ or a $k^+$- power is a word that is a $\beta$- power for some $\beta > k$.*

**Example 13**

1. *The German word schematische is a $\frac{3}{2}$- power of schemati and the English word entente is a $2^+$- power of ent.*
2. *Let $A = \{a, b, c\}$ and $w = aaaba = (aaab)^{\frac{5}{4}}$ and $u = abcabca$ is a $2^+$- power since there exist $\beta = \frac{7}{3} > 2$ which is a $\frac{7}{3}$- power of a word abc. i.e. $u = (abc)^{\frac{7}{3}}$.*

Note that: -An overlap can also be defined as $2^+$- power and a square (resp. cube) is a 2(resp.3)power.

**Definition 1.3.11** *If we can write $w = u^n v$ where $n \geq 1$ is an integer and $v$ is a prefix of $u$, then we say $\frac{|w|}{|u|}$ is an exponent of $w$ then $k = n + \frac{|v|}{|u|}$.*

**Example 14** *let $A = \{a, b, c\}$ and $w = aacbaacbaac = (aacb)^2 aac$ then $w$ is a $\frac{11}{4}$- power which is $\frac{11}{4} = 2 + \frac{3}{4}$ where aac is a prefix of aacb.*

# Chapter 2

# Repetition in Words

One of the most intensively studied topics of combinatorics of words is that repetition in words initiated by A.Thue at the beginning of this century. It focuses on the factor of words instead of words them selves. In this chapter we see the most types of repetitions of squares, cubes and overlap in infinite words and their repetition freeness(avoids repetition) in binary and ternary alphabets and we define abelian repetition and their avoidability over binary and ternary alphabets by using Dekking's construction[11].

**Definition 2.0.12** *A word $w$ is said to contain a repetition of order $k$ with a rational $k > 1$, if it contains a factor of the form*

$$u \in pref(v^{\omega}), with \ \ \frac{|u|}{|v|} = k$$

*In particular if $|u| = 2|v|$ and $z = u_1vvu_2$ with $u_1, u_2 \in A^*$, then $z$ contains a repetition of order 2. i.e. A square as a factor.*

**Example 15** *The word $w =$ **abcccacbacba** contains square, cube and overlap words as a factor. i.e $|w| = 12$, and it contains a repetition of order $3$, $ccc$ then $u = vvv$ where $v = c$ and $|v| = 1$.*

Special emphasis has been put to study repetition-free words. We define two different variants of this notion as follows.

**Definition 2.0.13** *Let $k > 1$ and $k \in \Re$ then we say a word $w \in A^{\infty}$ is:*

1. *$k$- free if it does not contain as a factor a repetition of order at least $k$.*

2. *$k^+$ - free if for any $\beta > k$ it is $\beta$- free.*

Thus 2-free, $2^+$-free and 3-free are called as square-free, overlap-free and cube-free respectively.

Note that:

1. The 2-freeness means that $w$ does not contain a square as a factor.

2. The $2^+$-freeness means that $w$ can contain a square, but no factor of the form $auaua$, with $u \in A^*$ and $a \in A$.

**Example 16** *$w = babaabaab$, the highest order of repetition is $2\frac{2}{3}$, hence it is not $2\frac{2}{3}$-free since it contains the factor $(aba)^{2\frac{2}{3}} = ababaab$.*
*Although $w$ does not contain a factor of the form $v^{2\frac{2}{5}}$ so, it is $2\frac{2}{5}$-free, but it is not $2\frac{2}{5}^{+}$-free since it contains a repetition of order $2\frac{2}{3} > 2\frac{2}{5}$.*

**Lemma 1** *Let $h$ be a uniform morphism, and let $k = \beta$ (resp., $k^+$) for a real number $\beta \geq 1$. If $w$ contains an $k$-power then so does $h(w)$*

**Proof 4** *Suppose $w$ contains a $k$-power. Then there exist words $s$,$s' \in A^+$ and $r$, $t \in A^*$ such that $w = rs^n s't$, where $s'$ is a nonempty prefix of $s$ and $n + \frac{|s'|}{|s|} \geq \beta$ (resp., $> \beta$). Then $h(w) = h(r)h(s)^n h(s')h(t)$. Then $h(w)$ contains $h(s)^n h(s')$, which is of exponent $\geq k$.*

**Example 17** *$h : \{a, b, c\}^* \to \{a, b, c\}^*$ be a morphism defined by:*

$$a \to abc, \quad b \to bac, \quad c \to acb$$

*Let $w = acacb$, then $w$ is a $2\frac{1}{2}$-power where $(ac)^2 b = (ac)^{\frac{5}{2}}$ then $h(w) = h(acacb) = abcacbabcacbbac$ which is also $2\frac{1}{2}$. i.e.$(abcacb)^2 bac = (abcacb)^{\frac{5}{2}}$*

**Thue's Problem. Find as long as possible, preferably infinite, word over an n-letter alphabet such that it is $k$-free (or $k^+$-free).**

Now we go to solutions of Thue's Problems.

First we characterize words in terms of how different occurrences of factors can situate inside a word. We note that two distinct occurrences of a factor $u$ of $w$, then there are words $x, y, x', y' \in A^*$ such that:

$$w = xuy = x'uy', x \neq x'$$

These two occurrences of $u$ either overlap or are consecutive or are disjoint. More precisely, we may suppose $|x| \leq |x'|$. Then three possibilities arise.

1. $|x'| > |xu|$. In this case, $x' = xuz$ for some $z \in A^+$ , and $w = xuzuy'$. The occurrences of $u$ are disjoint.
2. $|x'| = |xu|$. This implies that $x' = xu$, and consequently $w = xuuy'$ contains a square. The occurrences of $u$ are adjacent.
3. $|x'| < |xu|$. The two occurrences of $u$ are said to overlap. The following lemma gives a more precise description of this case.

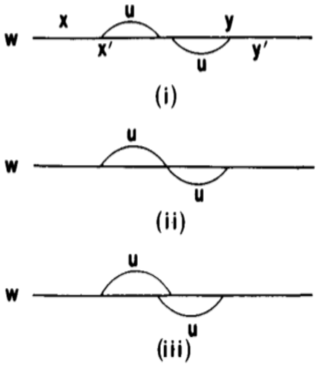
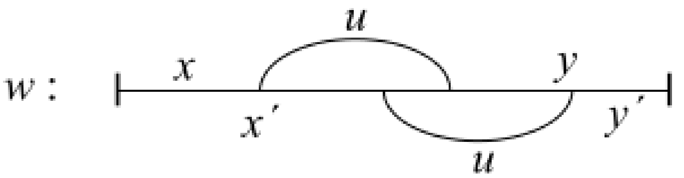


Figure 1. i.disjoint ii. consecutive iii. overlap

**Lemma 2** *Let $w \in A^\infty$, $w$ contains two overlapping occurrence of a word $u \neq \epsilon$ if and only if $w$ contains a factor of the form avava with $a \in A$, $v \in A^*$.*

**Proof 5** *(⇒) Assume $w$ contains two overlapping occurrences of a word $u$ and show that $w$ contains a factor of the form avava. We assume $w = xuy = x'uy'$ with $|x| < |x'| < |xu| < |x'u|$, we have an illustration:*



*Consequently,*

$x' = xs, \quad xu = x'z, \quad x'u = xut \ for \ some \ nonempty \ words \ s, \ z, \ t \ whence:$

$$u = sz = zt \tag{2.1}$$

*let* $a$ *the first letter* ($a = pref_1(s)$) *and therefore also for* $z$ *by* (2.1), *set* $s = av, \quad z = az'$ *then by* (2.1) $u = avaz'$ *and*

$$w = xsuy' = xavavaz'y'$$

*Hence* $w$ *contains a factor of the form* $avava$.
($\Leftarrow$) *if* $avava$ *is a factor of* $w$, *then* $u = ava$ *has two overlapping occurrence in* $w$.

Thus according to the lemma, a word has two overlapping occurences of a word if and only if it contains an overlapping factor.

**Theorem 2.0.1** *Let* $w \in A^\infty$. *Then the following conditions are equivalent:*

1. $w$ *is* $2^+$*-free;*
2. $w$ *does not contain a factor of the form* $avava$ *with* $a \in A; v \in A^*$.
3. $w$ *does not contain an overlapping factor.*

**Proof 6** (1) $\Rightarrow$ (2) *Assume* $w$ *is* $2^+$*-free. We want to show that* $w$ *does not contain a factor of the form* $avava$ *with* $a \in A, \ v \in A^*$.
*Assume that* $w$ *contains* $avava$ *as a factor, then it contains a repetition of order at least* $2 + \frac{1}{|v|+1}$, *contradiction to the assumption that* $w$ *is* $2^+$*-free. hence* $w$ *does not contain a factor of the form* $avava$ *with* $a \in A', \ v \in A^*$
(2) $\Rightarrow$ (3) *Assume* $w$ *does not contain a factor of the form* $avava$ *with* $a \in A, \ v \in A^*$. *We want to show that* $w$ *does not contain an overlapping factor. Assume that* $w$ *contains an overlapping factor* $u$, *then by lemma above it contains a factor of the form* $avava$ *with* $a \in A$ *and* $v \in A^*$. *contradiction since* $w$ *does not contains a factor of the form* $avava$ *with* $a \in A, \ v \in A^*$.
(3) $\Rightarrow$ (1) *Assume* $w$ *does not contain an overlapping factor. We want to show that* $w$ *is* $2^+$*-free. Now, assume* $w$ *is not* $2^+$*-free, then it contains a repetition of order* $k$, *with* $k > 2$, *and hence a factor* $(av)^2a$. *This means that* $w$ *contains an overlapping factor* $avava$ *where* $a = pref_1(u)$.

## 2.1 The Infinite Words of Thue-Morse

In this section a special infinite word is defined and its properties are studied. The main result is that this infinite word has no overlapping factor.

**Definition 2.1.1** *The Thue-Morse word* $\alpha = \alpha_0\alpha_1\cdots\alpha_n\cdots$ *is the binary word* $\alpha : \mathbb{N} \to \{0,1\}$ *defined recursively by* $\alpha_0 = 0$*, and for all* $n \geq 0$*,* $\alpha_{2n} = \alpha_n$ *and* $\alpha_{2n+1} = \overline{\alpha_n}$*, where* $\overline{a} = 1 - a$ *for* $a \in \{0,1\}$*.*

$$\alpha = \alpha_0\alpha_1\alpha_2\alpha_3\cdots\alpha_m\cdots\alpha_{2m}\alpha_{2m+1}\cdots$$

$$= 0 \;\; 1 \;\; 1 \;\; 0\cdots \; a\cdots \;\; a \;\; \overline{a}\cdots$$

Recall the Thue-Morse morphism $\mu : \{a,b\}^* \to \{a,b\}^*$ defined by:

$$a \to ab, \quad b \to ba$$

Consequently, iteration of $\mu$ on $a$ and on $b$ yields two infinite words

$$\alpha = \mu^\omega(a), \quad \overline{\alpha} = \mu^\omega(b)$$

then $\alpha$ is infinite word of Thue-Morse word. Computation gives

$$\begin{aligned} \mu(a) &= ab & \mu(b) &= ba \\ \mu^2(a) &= abba & \mu^2(b) &= baab \\ \mu^3(a) &= abbabaab & \mu^3(b) &= baabbabba \end{aligned}$$

$$\alpha = abbabaabbaababbabaababba\cdots$$

$$\overline{\alpha} = baababbaabbabaababba\cdots$$

Since $\mu$ is prolongabele on both letters, there are several property relating the word $\mu^n(a)$ and $\mu^n(b)$ for $n \geq 0$.
Consider the morphism $w \to \overline{w}$ defined by $\overline{a} = b$ and $\overline{b} = a$
Thus $\overline{w}$ obtained from $w$ by replacing each $a$ by $b$ and conversely. Of course $\overline{\overline{w}} = w$.

**Proposition 2** *Define* $u_o = a, v_o = b$ *and for* $n \geq 0$

$$u_{n+1} = u_nv_n, v_{n+1} = v_nu_n$$

*Then for all* $n \geq 0$

1. $u_n = \mu^n(a), \qquad v_n = \mu^n(b)$

2. $v_n = \overline{u}_n, \quad u_n = \overline{v}_n$.

3. $u_{2n}, v_{2n}$ *are palindromes and* $\tilde{u}_{2n+1} = v_{2n+1}$.

**Proof 7** *The proofs are by induction:*

1. *For* $n = 0$, $u_0 = a$ *and* $\mu^0(a) = a$ *and* $v_0 = b$ *and* $\mu^0(a) = a$.*For* $n = 1$, $u_1 = u_0v_0 = ab$ *and* $\mu^1(a) = ab$ *and* $v_1 = v_0u_0 = ba$ *and* $\mu^1(b) = ba$.
   *Assume it is true for* $n$ *i.e.*$u_n = \mu^n(a)$ *and* $v_n = \mu^n(b)$. *Then we want to proof it is true for* $n+1$.
   *Then,*
   $$u_{n+1} = u_nv_n = \mu^n(a)\mu^n(b) = \mu^n(ab) = \mu^n(\mu(a)) = \mu^{n+1}(a)$$
   $$v_{n+1} = v_nu_n = \mu^n(b)\mu^n(a) = \mu^n(ba) = \mu^n(\mu(b)) = \mu^{n+1}(b)$$
   *Next*

2. *follows from*
   $$v_{n+1} = v_nu_n = \overline{u}_n\overline{v}_n = \overline{u_nv_n} = \overline{u}_{n+1}, \overline{v}_{n+1} = \overline{\overline{u}}_{n+1} = u_{n+1}$$

3. $\tilde{u}_{2n} = u_{2n}$, *For* $n = 0, \tilde{u}_0 = a = u_0$, *for* $n = 1, \tilde{u}_2 = (u_1v_1)^{\sim} = \tilde{v}_1\tilde{u}_1 = u_1v_1 = u_2$
   *Assume it is true for* $n$, *i.e.* $\tilde{u}_{2n} = u_{2n}$, *then we want to proof for* $n+1$
   $$\tilde{u}_{2(n+1)} = \tilde{u}_{2n+2} = (u_{2n+1}v_{2n+1})^{\sim} = \tilde{v}_{2n+1}\tilde{u}_{2n+1} = u_{2n+2}$$
   *Hence* $u_{2n}$ *is a palindrome, simillarly we can proof for* $v_{2n}$.
   *Let* $k = 2n+1$ *and* $k > 0$, *then for* $k = 1, \tilde{u}_1 = (u_0v_0)^{\sim} = v_0u_0 = v_1$, *for* $k = 2, \tilde{u}_2 = (u_1v_1)^{\sim} = \tilde{v}_1\tilde{u}_1 = \tilde{v}_1v_1 = u_2$.
   *Assume it is true for* $k$, *i.e.* $\tilde{u}_k = (u_{k-1}v_{k-1})^{\sim} = \tilde{v}_{k-1}\tilde{u}_{k-1}$,
   *If* $k$ *is odd(resp. even) this implies*
   $$\tilde{u}_k = v_{k-1}u_{k-1} = v_k(resp.\ \ \tilde{u}_k = u_{k-1}v_{k-1} = u_k)$$
   *then we want to proof for* $k+1, \tilde{u}_{k+1} = (u_kv_k)^{\sim} = \tilde{v}_k\tilde{u}_k$
   *If* $k$ *is odd(resp. even), then* $k+1$ *is even(resp. odd) this implies*
   $$\tilde{u}_{k+1} = \tilde{v}_k\tilde{u}_k = u_kv_k = u_{k+1}(resp\ \ \tilde{u}_{k+1} = \tilde{v}_k\tilde{u}_k = v_ku_k = v_{k+1})$$

Our first characterization of the Thue-Morse word is in terms of binary expansions of nonnegative integers. First let $d_2(n)$ for $n \geq 0$ be the number of $1's$ in the binary expansion of $n$. For example the binary expansion of 9 is 1001, so $d_2(9) = 2$. Then we have the following proposition.

**Proposition 3** *For each* $n \geq 0$

$$\alpha_n = \begin{cases} a & \textit{if } d_2(n) \equiv 0 mod 2 \\ b & \textit{if } d_2(n) \equiv 1 mod 2 \end{cases} \tag{2.2}$$

**Proof 8** *By* (2.2) *we have*

$$\alpha = \mu(\alpha) = \mu(\alpha_0)\mu(\alpha_1)\cdots\mu(\alpha_n)\cdots$$

*and therefore* $\mu(\alpha_n) = \alpha_{2n}\alpha_{2n+1}$, *for* $n \geq 0$. *Then bythe definition of* $\mu$ *this implies*

$$\alpha_{2n} = \alpha_n, \quad \alpha_{2n+1} = \overline{\alpha}_n, \tag{2.3}$$

*Formula* (2.2) *holds for* $n = 0$. *Thus let* $n \geq 0$, *if* $n = 2m$, *then* $\alpha_n = \alpha_m$ *by* (2.3) *and* $d_2(n) = d_2(m)$. *Thus* (2.2) *holds in this case. If* $n = 2m + 1$, *then* $\alpha_n = \overline{\alpha_m}$ *and* $d_2(n) \equiv 1 + d_2(m) mod 2$. *There fore* (2.2) *holds in this case too.*

The inspection of $\alpha$ shows that $\alpha$ is not square-free. However, we will prove the second characterization of Thue-Morse word.

**Theorem 2.1.1** *There exists an infinite* $2^+$*-free word over a binary alphabet. In particular Thue-Morse word* $\alpha$ *is such.*

**Proof 9** *based on the two lemmas*

**Lemma 3** *Let* $X = \{ab, ba\}$. *If* $x \in X^*$, *then* $axa, bxb \notin X^*$

**Proof 10** *By induction on* $|x|$.

1. $|x| = 0$. *Clear since* $aa, bb \notin X^*$
2. *Assume that* $x \in X^*$ *with* $x \neq \epsilon$. *Assume further that* $u = axa \in X^*$ *(the case* $bxb \in X^*$ *being symmetric). We can write*

   $$u = abx_1 \cdots x_{r-1}ba \quad with \quad r \geq 1; x_i \in X$$

   *Set* $y = x_1 \cdots x_{r-1}$ *so that* $x = byb$ *with* $y \in X^*$. *Hence, by induction hypothesis* $x \notin X^*$, *a contradiction. So necessarily* $axa \notin X^*$

**Lemma 4** *Let* $w \in \{a, b\}^+$. *w contains an overlap if and only if* $\mu(w)$ *contains an overlap.*

**Proof 11** *($\Rightarrow$) Suppose $w$ contains an overlap, say $w = xcycycz$ for $x, y, z \in A^*$ and $c \in A$. Then*

$$\mu(w) = \mu(xcycycz) = \mu(x)\mu(c)\mu(y)\mu(c)\mu(y)\mu(c)\mu(z) = \mu(x)c\bar{c}\mu(y)c\bar{c}\mu(y)c\bar{c}\mu(z),$$

*and so, $\mu(w)$ contains an overlap $cvcvc$, where $v = \bar{c}\mu(y)$.*
*($\Leftarrow$) Assume that $\mu(w)$ has an overlapping factor for some word $w \in A^*$. We want to show that $w$ also has an overlapping factor.*
*By lemma 2 it can be assumed to be of the form $cvcvc$ with $c \in \{a, b\}$ and $v \in \{a, b\}^*$, that is:*
*we can write*

$$\mu(w) = xcvcvcy \quad where \quad x,\ v,\ y \in A^*$$

*Note that $|cvcvc|$ is odd, but $\mu(w) \in \{ab, ba\}^*$ with $X = \{ab, ba\}$, therefore $|\mu(w)|$ is even and $|xy|$ is odd. Thus*
*Either $|x|$ is even, and $x, cvcv, cy \in X^*$,*
*or $|x|$ is odd, and $xc, vcvc, y \in X^*$*
*This implis that $|v|$ is odd, since otherwise we get from $cvcv \in X^*$ (resp.$vcvc \in X^*$) that both $v, cvc$, are in $X^*$. Contradiction with the above lemma.*
*In case $|x|$ is even it follows that $cv \in X^*$ and $w = rsst$, where $\mu(r) = x, \mu(s) = cv$ and $\mu(t) = cy$. But then $s$ and $t$ start with letter $c$ and $ssc$ is an overlapping factor of $w$.*
*In case $|x|$ is odd, similarly $vc \in X^*$, and $w = rsst$, where $\mu(r) = xc, \mu(s) = vc$ and $\mu(t) = y$. Here $r$ and $s$ end with $c$, and $css$ is an overlapping factor of $w$.*

Now we proof Theorem 2.0.1

**Proof 12** *Assume that $\alpha$ is not $2^+$-free. Hence, by lemma 2, it has an overlapping factor. This means that for some $i \geq 0$, $\mu^i(a)$ has an overlapping factor as well. Then, by the above lemma, also $\mu^{i-1}(a)$, and hence inductively $a$ has an overlapping factor, a contradiction.*
*It follows that is $2^+$-free.*

**Corollary 1** *$\alpha$ is cube-free.*

**Proof 13** *Since $\alpha$ does not contain overlapping factor by the above theorem, then for any $\beta > 2$, $\alpha$ is $\beta$-free. Since $3 > 2$ $\alpha$ is cube free.*

Using the fact that $\alpha$ is overlap free, we may now construct a square-free infinite word over the alphabet of size 3 using Thue-Morse word[2].

## 2.2 Infinite Square-Free Words

Squares ca not be avoided in infinite binary words. For example the infinite word of Thue-Morse has square factors. In fact, the only square-free words over two letters $a$ and $b$ are:

$$a, b, ab, ba, aba, bab$$

**Theorem 2.2.1** *There is no square-free word over* $2$ *alphabets of length* $\geq 4$.

**Proof 14** *Assume* $w$ *is square-free and* $|w| \geq 4$. *Then with out loss of generality we may assume the first letter of* $w$ *is* $a$. *Then the second symbol must be* $b$*, for other wise we would has the square* $aa$. *Then the third symbol must be* $a$*,for other wise we would have the square* $bb$. *Thus the first 3 letters are* $aba$ *and what ever the letter we choose next gives a square.contradicts to our assumption.*

But we proof next there exist infinite square-free words over three letters. This will now be demonstrated.
Let $A = \{a, b\}$ and $B = \{a, b, c\}$. Define a morphism $\delta : B^* \to A^*$ by

$$a \to abb, \quad b \to ab, \quad c \to a$$

For any infinite word $\mathbf{b}$ on $B$,

$$\delta(\mathbf{b}) = \delta(b_0)\delta(b_1)\cdots\delta(b_n)\cdots$$

is an infinite word on $A$ starting with a letter $a$. Conversly consider an infinite word $\mathbf{a}$ on $A$ with out overlapping factor and starting with $a$. Then $\mathbf{a}$ can be factored as:

$$\mathbf{a} = y_0 y_1 \cdots y_n \cdots \tag{2.4}$$

with $y_n \in \{a, ab, abb\}$. Indeed each $a$ in $\mathbf{a}$ is followed by at most two $b$ since $bbb$ is an overlapping, and then followed by a new $a$. Moreover the factorization (2.4) is unique. Thus there exists a unique infinite word $\mathbf{b}$ on $B$ such that $\delta(\mathbf{b}) = \mathbf{a}$.

**Theorem 2.2.2** *Let* $\boldsymbol{a}$ *be an infinite word on* $A$ *starting with* $a$*, and without overlapping factor, and let* $\boldsymbol{b}$ *be the infinite word over* $B$ *such that* $\delta(\boldsymbol{b}) = \boldsymbol{a}$*; then* $\boldsymbol{b}$ *is square-free.*

**Proof 15** *Assume the contrary. Then* $\boldsymbol{b}$ *contains a square, say* $uu$. *Let* $d$ *be the letter following* $uu$ *in one of its occurrences in* $\boldsymbol{b}$. *Then* $\delta(uud)$ *is a factor of* $\boldsymbol{a}$. *Since* $\delta(u) = av$ *for some* $v \in A^*$ *and* $\delta(d)$ *starts with* $a$*,* $\boldsymbol{a}$ *contains the factor* $avava$. *Contradiction*

By applying the theorem to the Thue-Morse word $\alpha$, we obtain an infinite square-free word $\mathbf{m}$ over the three letter alphabet $B$ such that $\delta(\mathbf{m}) = \alpha$. This infinite word is:

$$\mathbf{m} = abcacbabcbacabcacbacabcbabcacbabcbacabcbab\cdots$$

**Theorem 2.2.3** *There exist an infinite* 2*-free word over a ternary alphabet.*

**Proof 16** *Let* $\boldsymbol{m}$ *be the word defined above. if it would contain a square, say* $uu$*, as a factor, then* $\delta(\boldsymbol{m}) = \alpha$ *would contain* $\delta(u)\delta(u)a$ *as a factor. This, however, is overlapping, by the form of* $\delta$*, proving the theorem.*

**OR**

Constructing a squarefree word from $\alpha$ as follows: Count the number of $b$'s in $\alpha$ between consecutive $a$'s.We get

$$a\,\overbrace{bb}^{2}\,a\,\overbrace{b}^{1}\,\overbrace{\quad}^{0}\cdots$$

is squarefree.

**Theorem 2.2.4** *For* $n \geq 1$ *define* $t_n$ *to be the number of* $b$*'s between the* $n^{th}$ *and* $(n+1)^{st}$ *occurance of* $a$ *in the word* $\alpha$*. Set* $t = t_1t_2\cdots$*. Then* $t = 21020121012\cdots$ *is an infinite squarefree word over the alphabet* $\{0, 1, 2\}$*.*

**Proof 17** $t$ *is over* 3 *alphabet*
*For if there were three or more* $b$*'s between two consecutive occurance of* $a$*'s in* $\alpha$*, then* $\alpha$ *would not be overlappfree.a contradiction. Next assume* $t$ *is not squarefree, then it contains a square of the form* $xx$*, with* $x = x_1x_2\cdots x_n$ *and* $n \geq 1$*. Then from the definition of* $t$*,* $\alpha$ *contains a factor of the form:*

$$ab^{x_1}ab^{x_2}a\cdots ab^{x_n}ab^{x_1}a\cdots ab^{x_n}a$$

*which contains an overlapp.a contradiction to* $\alpha$*.*
*Hence* $t$ *does not contain a square over ternary alphabet.*

Thue's Problem has natural commutative variants when $k \in \mathbb{N}$

## 2.3 Abelian Repetition

An abelian square is a word of the form $xx'$ with $|x| = |x'|$ and $x'$ is apermutation of $x$ and abelian cube is a word of the form $xx'x''$ with $|x| = |x'| = |x''|$ and $x', x''$ are both permutations of $x$. For example the English words reappear and deeded.
Similarly we can define abelian $k^{th}$-power for a positive integer $k \geq 2$.
Let $A = \{a_1, a_2, \cdots, a_k\}$, we use $\mathbb{Z}$ to denote the set of integer entries. For $u, v \in A^*$, we write $u \backsim v$ if $u$ and $v$ are anagrams of each other. i.e if $|u|_a = |v|_a$ for all $a \in A$. Then we define:

**Definition 2.3.1** *For an alphabet* $A = \{a_1, a_2, \cdots, a_k\}$, $\Pi : A^* \to \mathbb{Z}^k$ *is called a parikh map, if for all* $w \in A^*$, $\Pi(w) = [|w|_{a_1}, \cdots, |w|_{a_k}]$, *where* $|w|_{a_i}$ *is the number of occurrence* $a_i$ *in* $w$. *In other word,* $\Pi(w)$ *is a row vector which counts the occurrence of* $a_1, a_2, \cdots, a_k$ *in* $w$.

Note that: For $u, v \in A^*$ we have $u \backsim v$ exactly $\Pi(u) = \Pi(v)$.

**Definition 2.3.2** *Let* $k$ *be a positive integer, an abelian* $k$*-power is a nonempty word of the form* $w_1 w_2 \cdots w_k$ *where* $w_i \backsim w_{i+1}$, $1 \leq i \leq k-1$

**Definition 2.3.3** *We say that a word* $w$ *contains an abelian repetition of order* $k$, *if it contains a factor* $u_1 \cdots u_k$ *such that* $\Pi(u_1) = \Pi(u_2) = \cdots = \Pi(u_k)$, *where the mapping* $\Pi$ *gives the commutative image of a word, i.e*

$$\Pi(u) = (|u|_{a_1}, \cdots |u|_{a_n}), \quad \textit{when} \quad A = \{a_1 \cdots a_n\}$$

**Definition 2.3.4** *A word is abelian* $k$*-free if it does not contain an abelian repetition of order* $k$.

Recall the incidence matrix of a morphism:
Let $h : A^* \to B^*$ be a morphism for some finite set $A = a_1, a_2, \cdots a_d$, Then we define incident matrix $M = M(h)$ associated with the morphism $h$ is the matrix given by:

$$M = (m_{i,j})_{1 \leq i,j \leq d}$$

Where $m_{i,j}$ is the number of occurrences of $a_i$ in $h(a_j)$, i.e. $m_{i,j} = |h(a_j)|$ and
Let us also define the map $\psi : A^* \to \mathbb{Z}^d$ by $\psi(w) = [|w|_{a_1}, |w|_{a_2}, \cdots, |w|_{a_d}]^T$
The matrix $M(h)$ is useful because of the following proposition:

**Proposition 4**

$$\psi(h(w)) = M(h)\psi(w)$$

**Proof 18** $\psi(h(w)) = [|h(w)|_{a_1}, |h(w)|_{a_2}, \cdots, |h(w)|_{a_i}]^T$ *but we have*

$$|h(w)|_{a_i} = |h(a_1)|_{a_i}|w|_{a_1} + |h(a_2)|_{a_i}|w|_{a_2} + \cdots + |h(a_j)|_{a_i}|w|_{a_j}$$

$$= \sum_{1 \leq j \leq d} |h(a_j)|_{a_i}|w|_{a_j}$$

*Then*

$$\psi(h(w)) = \begin{pmatrix} |h(a_1)|_{a_1}|w|_{a_1} + |h(a_2)|_{a_1} + \cdots + |h(a_j)|_{a_1}|w|_{a_j} \\ \vdots \\ |h(a_1)|_{a_j}|w|_{a_1} + |h(a_2)|_{a_j} + \cdots + |h(a_j)|_{a_j}|w|_{a_j} \end{pmatrix}$$

$$= \begin{pmatrix} |h(a_1)|_{a_1} & |h(a_2)|_{a_1} & \cdots & |h(a_j)|_{a_1} \\ |h(a_1)|_{a_2} & |h(a_2)|_{a_2} & \cdots & |h(a_j)|_{a_2} \\ \vdots & \vdots & \cdots & \vdots \\ |h(a_1)|_{a_i} & |h(a_2)|_{a_i} & \cdots & |h(a_j)|_{a_i} \end{pmatrix} \times \begin{pmatrix} |w|_{a_1} \\ \vdots \\ |w|_{a_j} \end{pmatrix}$$

$$= M(h)\psi(w)$$

### 2.3.1 Dekking Construction

In this section we explore a construction due to Dekking [4,7] that gives optimal results for abelian-power-free words over alphabets of size 2 and 3. We start with some definitions about morphisms and groups.

**Definition 2.3.5** *Let* $h : A^* \to B^*$ *be a morphism. If* $w = h(a)$ *is the image of a single letter* $a \in A$ *we call it a****block****. If* $h(a) = vv', v \neq \epsilon$*, then we call* $v$ *a* ***left subblock*** *of* $v'$ *its corresponding* ***right subblock****.*

**Definition 2.3.6** *Let* $G$ *be a* ***finite abelian group*** *(written additively). We say that a subset* $A \subseteq G$ *is* ***progression-free*** *of order* $n$ *if for all* $a \in A$*,* $a, a+g, a+2g, \cdots a+(n-1)g \in A$ *implies that* $g = 0$*.*

**Example 18** *Let* $G = \mathbb{Z}/(7)$*, integer modulo 7, and let* $A = \{0, 1, 2, 4\}$*. Then* $A$ *is progression-free of order* 4*. For example, for each* $a \in A$*, the following table shows that for each* $g \neq 0$ *there exists* $i$*,* $0 \leq i \leq 3$ *such that* $a + ig \notin A$*.*

| a | g | i | a+ig |
|---|---|---|---|
| 0 | 1 | 3 | 3 |
| 0 | 2 | 3 | 6 |
| 0 | 3 | 1 | 3 |
| 0 | 4 | 3 | 5 |
| 0 | 5 | 1 | 5 |
| 0 | 6 | 1 | 6 |

| a | g | i | a+ig |
|---|---|---|---|
| 1 | 1 | 2 | 3 |
| 1 | 2 | 1 | 3 |
| 1 | 3 | 3 | 3 |
| 1 | 4 | 1 | 5 |
| 1 | 5 | 1 | 6 |
| 1 | 6 | 2 | 6 |

| a | g | i | a+ig |
|---|---|---|---|
| 2 | 1 | 1 | 3 |
| 2 | 2 | 2 | 6 |
| 2 | 3 | 1 | 5 |
| 2 | 4 | 1 | 6 |
| 2 | 5 | 2 | 5 |
| 2 | 6 | 3 | 6 |

| a | g | i | a+ig |
|---|---|---|---|
| 4 | 1 | 1 | 5 |
| 4 | 2 | 1 | 6 |
| 4 | 3 | 2 | 3 |
| 4 | 4 | 2 | 5 |
| 4 | 5 | 3 | 5 |
| 4 | 6 | 1 | 3 |

**Definition 2.3.7** *Let* $f = A^* \to G$ *be a morphism, so that* $f(\epsilon) = 0$ *the identity element of* $G$*, and* $f(a_1 a_2 \cdots a_i) = \sum_{1 \leq j \leq i} f(a_j)$*. We call* $f$ *is h-injective if for any collection* $v_1, v_2 \cdots v_n$ *of left subblocks and* $v'_1, v'_2 \cdots v'_n$ *the corresponding right subblocks, the equality* $f(v_1) = f(v_2) = \cdots = f(v_n)$ *or* $f(v'_1) = f(v'_2) = \cdots = f(v'_n)$ *implies that either* $v_1 = v_2 = \cdots = v_n$ *or* $v'_1 = v'_2 = \cdots = v'_n$*.*

**Lemma 5** *Let $n$ be a positive integer, let $h : A^* \to A^*$ be a morphism, let $G$ be a finite abelian group, and let $f : A^* \to G$ be a morphism such that:*

1. *The adjacency matrix of $h$ has nonzero determinant;*
2. *$f(h(a)) = 0$ for all $a \in A$.*
3. *The set $A = \{g \in G : g = f(v), v$ a left subblock of $h\}$ is progression-free of order $n + 1$.*
4. *$f$ is $h$-injective.*

*If $h$ is prolongable on $a$, and $h^\omega(a)$ avoids abelian $n^{th}$ powers $x_1x_2\cdots x_n$ where $|x_i| \leq max_{a\in A}|h(a)|$, then $h^\omega(a)$ is abelian $n^{th}$ power-free.*

**Proof 19** *Let $\boldsymbol{x} = h^\omega(a)$. By the hypothesis, $\boldsymbol{x}$ avoids "short" abelian n'th powers, that is, factors of the form $x_1x_2\cdots x_n$ where each $x_i$ is a permutation of $x_1$ and $|x_i| \leq max_{a\in A}|h(a)|$.*
*Suppose $B_1B_2\cdots B_n$ is an abelian n'th-power occurring in $\boldsymbol{x}$, with*

$$|B_1| = |B_2| = \cdots = |B_n|$$

*and each $B_i$ a permutation of $B_1$, and $|B_i|$ is minimal.*
*Since we have ruled out short powers, we must have $|B_i| > max_{a\in A}|h(a)|$.*
*Consider the factorization of $\boldsymbol{x}$ into blocks, each an image of a letter under $h$. Then each $B_i$ starts inside some block $h(a)$; let $v_i$ be the corresponding left subblock and $v'_i$ the corresponding right subblock, so that $h(a) = v_iv'_i$, and $B_i$ occurs starting at the same position where $v'_i$ starts, and $B_n$ ends at the same position where $v_{n+1}$ ends. (We take $v_i = \epsilon$ if a $B_i$ occurs starting at the same position as the beginning of a block.) By the length condition on the $B_i$, each $B_i$ starts in a distinct block. See Figure below, where this is illustrated for $n$ = 3.*

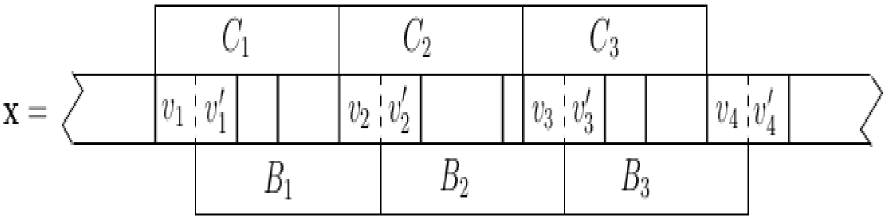


*figure 2. An abelian cube and its coresponding blocks Since each* $B_i$ *is a permutation of every other* $B_i$*, it follows that each* $B_i$ *contains exactly the same number of every letter, and so*

$$f(B_1) = f(B_2) = \cdots = f(B_n) \tag{2.5}$$

*On the other hand, from condition (b) of the Lemma, we know that* $f(h(a)) = 0$ *for every* $a \in A$*. Writing* $B_i = v'_i y_i v_{i+1}$ *for some* $y_i$ *that is the image of a word under* $h$*, we get* $f(B_i) = f(v'_i) + f(v_{i+1})$*. Since* $f(v_i v'_i) = 0$*, we get* $f(B_i) = -f(v_i) + f(v_{i+1})$*. From equation (2.5) we get that the* $f(v_i)$ *form an (n + 1)-term arithmetic progression with difference* $f(B_i)$*. But then by hypothesis (c), we get* $f(v_1) = f(v_2) = \cdots = f(v_{n+1})$*. Hence by hypothesis (d), it follows that either* $v_1 = v_2 = \cdots = v_{n+1}$ *or* $v'_1 = v'_2 = \cdots = v_{n+1}$*. In the former case, we can "slide" the abelian n'th power to the left by* $|v_1|$ *symbols and still get an abelian n'th power, in the latter case we can slide it to the right by* $|v_1|$ *symbols and still get an abelian n'th power. Now our abelian n'th power is aligned at both ends with blocks of h, so there is an abelian n'th power* $C_1C_2\cdots C_n$ *where each* $C_i$ *is composed of blocks, again see Figure above. Let* $D_i$ *be such that* $C_i = h(D_i)$*. Since* $\boldsymbol{x} = h(\boldsymbol{x})$*, it follows that* $D_1D_2\cdots D_n$ *occurs in* $\boldsymbol{x}$*. Now* $\psi(C_i) = M\psi(D_i)$*, where* $M$ *is the matrix of* $h$ *Since* $M$ *is invertible, there is only one possibility for* $\psi(D_i)$*. Since* $\psi(C_1) = \psi(C_2) = \cdots = \psi(C_n)$*, it follows that* $\psi(D_1) = \psi(D_2) = \cdots = \psi(D_n)$*. Hence* $D_1\cdots D_n$ *is a shorter abelian n'th power, contradicting the minimality of* $B_1B_2\cdots B_n$*.*

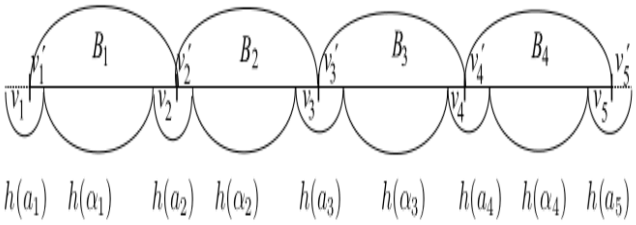
**Theorem 2.3.1** *There exist an infinite Abelian* 4*-free word over a binary alphabet.*

**Proof 20** *We shall prove that the infinite word* $\omega$ *defined by iterating the morphism*

$$h : \{a, b\}^* \to \{a, b\}^* \tag{2.6}$$

*defined by* $h(a) = abb$, $h(b) = aaab$ *at point* $a$ *is Abelian 4-free simply by showing that from a factor of* $\omega$ *which is Abelian 4th power we can construct a shorter similar factor. The fact that the consecutive blocks in the 4th power are only commutatively equal, and not equal, makes the proof much more complicated. As we shall see there are four, from (a) to (d), special properties of* $h$ *which are used in the following considerations.*
*First we associate with a word* $w$ *its value in the group* $\mathbb{Z}_5$ *by a morphism* $f : \{a, b\}^* \to \mathbb{Z}_5$ *defined as* $f(a) = 1$ *and* $f(b) = 2$*. This implies our first requirement for* $h$ *namely that*

1. $f(h(w)) = 0$ for all $w \in \{a,b\}^*$, which indeed is satisfied by our morphism (2.6). Now, assume that $B_1B_2B_3B_4$ is an Abelian 4-repetition in $\omega$. This together with the fact that these $B_i$'s are covered by h-images is illustrated as follows:

   Formally the above means that:

   $$h(a_1\alpha_1 \cdots \alpha_4 a_5) = v_1B_1B_2B_3B_4v_5' \quad \text{with} \quad a_i \in A \quad \text{and} \quad \alpha_j \in A^*$$

   where for $i = 1, \cdots 5$ and $j = 1, \cdots 4$

   $$h(a_i) = v_iv_i' \quad \text{and} \quad B_j = v_j'h(\alpha_j)v_{j+1} \quad \text{with} \quad v_i \in A^* \quad \text{and} \quad v_i' \in A^+$$

   Recall now that $f$ is a morphism so that, by (1), we obtain

   $$f(v_{j+1}) = f(B_j) - f(h(\alpha_j)) - f(v_j') = f(B_j) + f(v_j) = g + f(v_j)$$

   where $g$ is a constant since $B_j's$ are commutatively equal. It follows that the sequence

   $$f(v_1), f(v_2), f(v_3), f(v_4), f(v_5) \tag{2.7}$$

   is an arithmetic progression of length 5. This guides us to require that

2. $S = \{a \in \mathbb{Z}_5 \diagup \exists z \in pref\{h(a), h(b)\} : a = f(z)\}$ is 5-progression free, i.e. does not contain an arithmetic progression of length 5, with $g \neq 0$. That our morphism satisfies

   $$\{f(a), f(ab)\} = \{1,3\} \quad \text{and} \quad \{f(a), f(aa), f(aaa)\} = \{1,2,3\} \tag{2.8}$$

   so that $S = \{0,1,2,3\}$, while in $\mathbb{Z}_5$ any arithmetic progression of length 5, with $g \neq 0$, equals to the whole $\mathbb{Z}_5$.
   Since $v_i's$ in (2.7) are prefixes of $h(a)$ and $h(b)$ we can write (2.7) in the form

   $$f(v_1) = f(v_2) = f(v_3) = f(v_4) = f(v_5)$$

   We now return to the Figure . We want that $B_i$'s, possibly after a shift, would match with the $\{h(a), h(b)\}$-factorization of $\omega$. This is achieved if either the words $v_i$ or the words $v_i'$ coincide. This motivates our next condition required for $h$ and $f$. We say that $f$ is h-injective, if for all factorizations $v_iv_i' \in \{f(a), f(ab)\}$ with $i = 1, \cdots 5$, we have

3. $f(v_1) = f(v_2) = f(v_3) = f(v_4) = f(v_5) \Rightarrow v_1 = v_2 = \cdots v_5$ or $v_1' = v_2' = \cdots v_5'$ From our computations in (2.8) we see that the only case to be checked here is the case when $v_1 = ab$ and $v_2 = aaa$, and then indeed $v_1' = b = v_2'$. So for our choice of $f$ and $h$, $f$ is $h$-injective. We are almost done. The words $v_i$ or $v_i'$ coincide. Consequently, the four Abelian repetitions $B_i$ can be shifted to match with the morphism $h$: instead of $B_i's$ we now consider the commutatively equal blocks.

   $$D_i = v_i B_i v_1^{-1} (or D_i = v_i'^{-1} B_i v_i') \quad for \quad i = 1, \cdots 4$$

   Then there are words $C_i$ such that

   $$h(C_i) = D_i \quad with \quad \psi(D_i) = \psi(D_j) \quad for \quad i, j = 1, \cdots 4 \tag{2.9}$$

   where $\psi$ gives the commutative image of a word. If we would know that $C_i's$ were commutatively equal the proof would be complete. Indeed, then $\omega$ would contain a shorter Abelian 4-repetition, and hence inductively also either aaaa or bbbb as a factor. This, however, is not the case. So to complete the proof we impose one more requirement for h, namely that

4. $$M(h) = \begin{pmatrix} |h(a)|_a & |h(b)|_a \\ |h(a)|_b & |h(b)|_b \end{pmatrix} \quad is \quad invertible$$

   Then, by (2.9), we have

   $$\psi(C_i).M(h) = \psi(D_i) \quad for \quad i = 1, \cdots 4$$

   or equivalently that

   $$\psi(C_i) = \psi(D_i).M(h)^{-1} \quad for \quad i = 1, \cdot 4$$

   Our invertible matrix of $h$ is

   $$\mathbf{M(h)} = \begin{pmatrix} 1 & 3 \\ 2 & 1 \end{pmatrix}$$

   Which has determinant -5. Choose $G = \mathbb{Z}/(5)$ and define $f$ by $f(a) = 1, f(b) = 2$. Then $f(h(a)) = 0$ for $a \in \{a, b\}$.
   Furthermore $A = \{0, 1, 2, 3, \}$ which is progression free of order 5.
   Thus $f^{\omega}(0)$ is abelian $4^{th}$-power-free.
   We can check that there are no abelian $4^{th}$ powers $x_1x_2x_3x_4$ in $h^{\omega}(a)$ for $|x_1| \leq 4$ by enumerating all factors of length $\leq 16$.

**Theorem 2.3.2** *There exist an infinite Abelian cube-free word over a ternary alphabet.*

**Proof 21** *Let* $A = \{a, b, c\}$ *and define* $h$ *by* $h(a) = aabc, h(b) = bbc, h(c) = acc$
*Then the matrix of* $h$*:*

$$\begin{pmatrix} 2 & 0 & 1 \\ 1 & 2 & 0 \\ 1 & 1 & 2 \end{pmatrix}$$

*Which has determinant* $7$*.let* $G = \mathbb{Z}\diagup(7)$ *and define* $f(a) = 1, f(b) = 2,$ *and* $f(c) = 3$*. Then* $f(h(a)) = 0$ *for each* $a \in \{a, b, c\}$*. Then* $A = \{0, 1, 2, 4\}$ *which is progression free of order* $4$*.*

# Reference

1. M.Lothaire, *Combinatorics On Words*: book.org, 2004.
2. M.Lothaire, *Applied Combinatorics On Words*: Cambridge University Press, 1983.
3. Jean-Paul Allouche, Jeffrey Shallit, *Automatic sequences Theory, Application, Generalizations*: Cambridge University Press, 2003.
4. Narad Rampersad and Jeffery Shallit, *Repetition in words*: May, 2012.
5. Narad Rampersad, *Overlap-free words and Generalization*: Waterloo, Ontario, Canada, 2007.
6. A. Thue, Uber unendliche Zeichenreihen, Norske Vid. Selsk. Skr. I. Mat. Nat. Kl. Christiana 7 (1906) 1-22.
7. F. M. Dekking. Strongly non-repetitive sequences and progression-free sets, J. Comb. Theory Ser. A 27 (1979), 181-185.