On arithmetic index in the Thue-Morse word

Olga Parshina

Sobolev Institute of mathematics SB RAS
Institut Camille Jordan UCBL1

parolja@gmail.com

November 29, 2016

- Let $\Sigma_t = \{0, 1, ..., t-1\}$ be a finite alphabet
  A sequence $w = w_0 w_1 w_2 w_3 \cdots$, $w_i \in \Sigma_t$ is an infinite word over Σ_t
- A sequence u is *periodic* with the length of a period T,
  if for every i the following holds: $u_i = u_{i+T}$
- An *arithmetic subsequence* of length k with the starting symbol w_c
  and difference d in w is a word $w_d^c = w_c w_{c+d} w_{c+2d} \cdots w_{c+(k-1)d}$
- If $w_{c+id} = a$ for $i = 0, 1, ..., k-1$ and $a \in \Sigma_t$,
  then w_d^c is called an *arithmetic progression*.

- An *arithmetic closure* of an infinite word w over Σ_t is a set $A_w = \{w_d^c | c \geq 0, d > 0\}$
- An *arithmetic complexity* of w is a function $a_w(n) = |A_w \bigcap \Sigma_t^n|$
- A word w is *arithmetic universal*, if $a_w(n) = t^n$

Theorem [Van der Waerden, 1927]

An arithmetic closure of every infinite word over a finite alphabet contains an arbitrary long arithmetic progression.

Object of the research

Let $\Sigma = \{0, 1\}$, $x \in \mathbb{N}$

- $B(x) = x_{n-1}...x_1x_0$, where $x_i \in \Sigma, x = \sum_{i=0}^{n-1} x_i \cdot 2^i$ is a *binary expansion of x*
- $p(x) = \sum_{i=0}^{n-1} x_i \bmod 2$

The *Thue-Morse word* is a sequence $w_{TM} = w_0 w_1 w_2 w_3 \cdots$, where $w_i = p(i)$.

$$w_{TM} = 0110100110010110100101100110100 1 \cdots$$

Words with $T = 1$

Let $\Sigma = \{0, 1\}$, $w_{TM} = w_0 w_1 w_2 w_3 \cdots$ is the Thue-Morse word.

- Define $L(c, d)$ as the length of an arithmetic progression in w_{TM} with the starting symbol w_c and the difference d for $n, d \in \mathbb{N}$
- $L(d) = \max_c L(c, d)$

Theorem [P., 2015]

For every $n \in \mathbb{N}, n > 1$ the following holds:

$$\max_{d<2^n} L(d) = \begin{cases} 2^n + 4, & n \text{ is even,} \\ 2^n, & n \text{ is odd.} \end{cases}$$

Moreover, the maximum is reached with the difference $d = 2^n - 1$ in both cases.

Words with $T > 1$

Consider a word $u = (01)^k, k \in \mathbb{N}$
$u \in A_{TM}$

[Avgustinovich, Frid, Fon-der-Flaas, 2000]

A word w_{TM} is arithmetic universal.

Words with $T > 1$

Consider a word $u = (01)^k, k \in \mathbb{N}$
$u \in A_{TM}$

[Avgustinovich, Frid, Fon-der-Flaas, 2000]

A word w_{TM} is arithmetic universal.

Questions:

- What is an upper bound on the value of the difference while searching arithmetic subwords in the Thue-Morse word?
- Which words occur in the Thue-Morse word with the maximal difference?

Object of the research

- Let w be an infinite word over Σ_t.
- Let $d \in \mathbb{N}$, then define $A_w(d)$ as a set of arithmetic subsequences with the difference d in w.
- $A_w = \bigcup_{d=1}^{\infty} A_w(d)$
- A finite word u has an *arithmetic index* $I_w(u)$, if $I_w(u)$ is the length of a binary expansion of $i(u) = \min\{d | u \in A_w(d)\}$
- If u does not belong to A_w, then $I_w(u) = \infty$
- The object of a research is a function $\max_{u:|u|=n} I_w(u)$

Words with $T = 1$

- Define $L(c, d)$ to be the length of an arithmetic progression in w_{TM} with the starting symbol w_c and the difference d for $c, d \in \mathbb{N}$
- $L(d) = \max_c L(c, d)$

Theorem [P., 2015]

For every $n \in \mathbb{N}, n > 1$ the following holds:

$$\max_{d<2^n} L(d) = \begin{cases} 2^n + 4, & n \text{ is even}, \\ 2^n, & n \text{ is odd}. \end{cases}$$

Moreover, the maximum is reached with the difference $d = 2^n - 1$ in both cases.

Corollary

Let u be of the form 0^N or 1^N for some $N \in \mathbb{N}$.
Then $I_{TM}(u) = n = \lceil \log N \rceil$.

Upper bound

Let u be a binary word of length N, $2^{n-1} \leq N < 2^n$.

- Define a "basis": $b_1, b_2, ..., b_N$,
  $b_1 = 1\alpha_1\alpha_2\cdots\alpha_{N-1}$
  $b_2 = 01\alpha_1\alpha_2\cdots\alpha_{N-2}$
  $\cdots\cdots\cdots$
  $b_i = 0^{i-1}1\alpha_1\alpha_2\cdots\alpha_{N-i}$
  $\cdots\cdots\cdots$
  $b_N = 0^{N-1}1$, where $\alpha_i \in \{0,1\}$
- Every b_i can be obtained with the difference $d = 2^n - 1$
- A word u can be represented in the following way:
  $u = \sum_{i=1}^{n} \beta_i \cdot b_i$, $\beta_i \in \{0,1\}$

Upper bound

- Basis: $b_1, b_2, ..., b_N$,
  $b_i = 0^{i-1}1\alpha_1\alpha_2\cdots\alpha_{n-i}$, $\alpha_i \in \{0,1\}$
- $u = b_1 + b_2 + ... + b_N$

- Set C_i to be an index of the first symbol of b_i arithmetic occurrence in the Thue-Morse word, c_i to be a binary expansion of C_i, the length of each c_i is equal to $3n = 3 \cdot \lceil logN \rceil$
- Build a binary number $c = c_1 c_2 \cdots c_N$ of length $3n \cdot N$
- Then take a difference of the form
  $D = \underbrace{0\cdots0}_{2n}\underbrace{1\cdots1}_{n}\underbrace{0\cdots0}_{2n}\underbrace{1\cdots1}_{n}\cdots\underbrace{0\cdots0}_{2n}\underbrace{1\cdots1}_{n}$
- $\max_{u:|u|=N} l_w(u) \leq 3Nn = 3N \cdot \lceil logN \rceil$

Upper bound. Example

Set $u = 1110$, $N = 4 = 2^2, n = 2$

Define a basis:

- $d = 2^2 - 1 = 3$

$$\mathbf{w_{TM}} = \mathbf{0110100110010110100101100110 1001}\cdots$$

$b_1 = 1000$, $C_1 = 21$, $c_1 = 010101$
$b_2 = 0100$, $C_2 = 18$, $c_2 = 010010$
$b_3 = 0010$, $C_3 = 15$, $c_3 = 001111$
$b_4 = 0001$, $C_4 = 12$, $c_4 = 001100$

- $u = b_1 + b_2 + b_3$

Upper bound. Example

Set $u = 1110$, $N = 4 = 2^2, n = 2$

Define a basis:

- $d = 2^2 - 1 = 3$

$$\mathbf{w_{TM}} = \mathbf{01101001100101101001011001101001} \cdots$$

$b_1 = 1000$, $C_1 = 21$, $c_1 = 010101$
$b_2 = 0100$, $C_2 = 18$, $c_2 = 010010$
$b_3 = 0010$, $C_3 = 15$, $c_3 = 001111$
$b_4 = 0001$, $C_4 = 12$, $c_4 = 001100$

- $u = b_1 + b_2 + b_3$

Upper bound. Example

Set $u = 1110$, $N = 4 = 2^2, n = 2$

Define a basis:

- $d = 2^2 - 1 = 3$

$$\mathbf{w_{TM}} = \mathbf{0110100110010110100101100110100 1} \cdots$$

$b_1 = 1000$, $C_1 = 21$, $c_1 = 010101$
$b_2 = 0100$, $C_2 = 18$, $c_2 = 010010$
$b_3 = 0010$, $C_3 = 15$, $c_3 = 001111$
$b_4 = 0001$, $C_4 = 12$, $c_4 = 001100$

- $u = b_1 + b_2 + b_3$

Upper bound. Example

Set $u = 1110$, $N = 4 = 2^2, n = 2$

Define a basis:

- $d = 2^2 - 1 = 3$

$$\mathbf{w_{TM}} = \mathbf{011010011001011{\color{red}0}10{\color{red}0}10{\color{red}1}10{\color{red}0}1101001}\cdots$$

$b_1 = 1000,\ C_1 = 21,\ c_1 = 010101$
$b_2 = 0100,\ C_2 = 18,\ c_2 = 010010$
$b_3 = 0010,\ C_3 = 15,\ c_3 = 001111$
$b_4 = 0001,\ C_4 = 12,\ c_4 = 001100$

- $u = b_1 + b_2 + b_3$

Upper bound. Example

Set $u = 1110$, $N = 4 = 2^2, n = 2$

Define a basis:

- $d = 2^2 - 1 = 3$

$$\mathbf{w_{TM}} = \mathbf{01101001100101101001011001101001}\cdots$$

$b_1 = 1000$, $C_1 = 21$, $c_1 = 010101$
$b_2 = 0100$, $C_2 = 18$, $c_2 = 010010$
$b_3 = 0010$, $C_3 = 15$, $c_3 = 001111$
$b_4 = 0001$, $C_4 = 12$, $c_4 = 001100$

- $u = b_1 + b_2 + b_3$

Upper bound. Example

Set $u = 1110$, $N = 4 = 2^2, n = 2$

Define a basis:

- $d = 2^2 - 1 = 3$

$$\mathbf{w_{TM}} = \mathbf{0110100110010110100101100110100 1}\cdots$$

$b_1 = 1000$, $C_1 = 21$, $c_1 = 010101$
$b_2 = 0100$, $C_2 = 18$, $c_2 = 010010$
$b_3 = 0010$, $C_3 = 15$, $c_3 = 001111$
$b_4 = 0001$, $C_4 = 12$, $c_4 = 001100$

- $u = b_1 + b_2 + b_3$

Upper bound. Example

- Basis:
  $b_1 = 1000,\ C_1 = 21,\ c_1 = 010101$
  $b_2 = 0100,\ C_2 = 18,\ c_2 = 010010$
  $b_3 = 0010,\ C_3 = 15,\ c_3 = 001111$
  $b_4 = 0001,\ C_4 = 12,\ c_4 = 001100$
- $u = b_1 + b_2 + b_3$
- The initial number:
  $c = c_1 c_2 c_3 = 010101'010010'001111$
- The difference D:
  $D = 000011'000011'000011$
- We can obtain u as an arithmetic subword starting with w_C with the difference D:

Upper bound. Example

	c+iD	wt	u
	$010101'010010'001111$	9	1
$\oplus$	$000011'000011'000011$		
	$011000'010101'010010$	7	1
$\oplus$	$000011'000011'000011$		
	$011011'011000'010101$	9	1
$\oplus$	$000011'000011'000011$		
	$011110'011011'011000$	13	0

Lower bound

A *factor complexity* $p_w(N)$ of an infinite word w is the number of its factors of length N.

Theorem [Avgustinovich, 1994]

A factor complexity of the Thue-Morse word is
$p_{w_{TM}}(N+1) = 3N + \rho(N)$,
where $\rho(N) = min\{N - 2^k, 2^k + 1 - N\}$ and $n/2 < 2^k \leq N$.

- $|A_{TM}(1) \bigcap \Sigma^N| = p_{w_{TM}}(N) < 4N$

Lower bound

- Consider $A(d, N) := |A_{TM}(d) \bigcap \Sigma^N|$

- $A(1, N) \leq p_{w_{TM}}(N) < 4N$
- $A(3, N) \leq p_{w_{TM}}(3N) < 12N$

  $\cdots\cdots\cdots$
- $A(d, N) \leq p_{w_{TM}}(d \cdot N) < 4d \cdot N$

The aim is to obtain a lower bound for x s.t. $2^N \leq \sum_{d=1}^{x} A(d, N)$.

$$x \geq 0.5(N - 2 - \log N)$$

Resume

A word u has an arithmetic index $I_{TM}(u)$ in the Thue-Morse word,
if $I_{TM}(u)$ is the length of binary expansion of $i_{TM}(u) = \min\{d | u \in A_{TM}(d)\}$

$$0.5(N - 2 - \log N) < \max_{u:|u|=N} I_w(u) \leq 3N\lceil \log N \rceil$$

Thank you for your attention!