A conjecture on the 2-abelian complexity of the Thue-Morse word (Work in progress)

Aline Parreau, Michel Rigo, Elise Vandomme

Representing streams II 2014 – Leiden
The Thue-Morse word $\mathbf{t} = t_0 t_1 t_2 \cdots$ is the infinite word $\lim_{n \to +\infty} \varphi^n(0)$ where

$$\varphi : 0 \mapsto 01, \quad 1 \mapsto 10,$$

$$\mathbf{t} = 01101001100101101001011001101001 \cdots$$

The Thue-Morse word $\mathbf{t}$ is 2-automatic.

$$
\begin{array}{c|cccccccccc}
 n & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & \ldots \\
\hline
\text{rep}_2(n) & 0 & 1 & 10 & 11 & 100 & 101 & 110 & 111 & 1000 & \ldots \\
\hline
\mathbf{t}_n & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & \ldots \\
\end{array}
$$

\begin{center}
\begin{tikzpicture}
\node[state] (0) at (0,0) {$0$};
\node[state] (1) at (1,0) {$1$};
\node[state] (2) at (2,0) {$1$};
\node[state] (3) at (1,-1) {$0$};
\node[state] (4) at (0,-1) {$1$};
\node[state] (5) at (2,-1) {$1$};
\draw[->] (0) to [out=225,in=270,loop] node[below] {$0$} (0);
\draw[->] (0) to [out=315,in=90,loop] node[above] {$1$} (0);
\draw[->] (0) to node[above] {$1$} (1);
\draw[->] (1) to node[below] {$0$} (3);
\draw[->] (3) to node[below] {$1$} (2);
\draw[->] (2) to node[above] {$0$} (5);
\end{tikzpicture}
\end{center}
The factor complexity of the Thue-Morse word

\[ p_t(n) = \# \{ \text{factors of length } n \text{ of } t \} \]

is well-known: \( p_t(0) = 1, \quad p_t(1) = 2, \quad p_t(2) = 4, \)

\[
p_t(n) = \begin{cases} 
4n - 2 \cdot 2^m - 4 & \text{if } 2 \cdot 2^m < n \leq 3 \cdot 2^m, \\
2n + 4 \cdot 2^m - 2 & \text{if } 3 \cdot 2^m < n \leq 4 \cdot 2^m.
\end{cases}
\]

S. Brlek, Enumeration of factors in the Thue-Morse word, DAM’89
A. de Luca, S. Varricchio, On the factors of the Thue-Morse word on three symbols, IPL’88
Thue-Morse word

The factor complexity of the Thue-Morse word

\[ p_t(n) = \# \{ \text{factors of length } n \text{ of } t \} \]

is well-known: \( p_t(0) = 1, \ p_t(1) = 2, \ p_t(2) = 4, \)

\[ p_t(n) = \begin{cases} 4n - 2 \cdot 2^m - 4 & \text{if } 2 \cdot 2^m < n \leq 3 \cdot 2^m, \\ 2n + 4 \cdot 2^m - 2 & \text{if } 3 \cdot 2^m < n \leq 4 \cdot 2^m. \end{cases} \]

Definition

Two words \( u \) and \( v \) are abelian equivalent if \( |u|_\sigma = |v|_\sigma \) for any letter \( \sigma \).

The abelian complexity of \( t \) takes only two values

\[ P_t(2n) = 3 \text{ and } P_t(2n + 1) = 2. \]
Let $k \geq 1$ be an integer. Two words $u$ and $v$ in $A^+$ are $k$-abelian equivalent, denoted by $u \equiv_k v$, if

- $\text{pref}_{k-1}(u) = \text{pref}_{k-1}(v)$,
- $\text{suf}_{k-1}(u) = \text{suf}_{k-1}(v)$,
- for all $w \in A^k$, the number of occurrences of $w$ in $u$ and in $v$ coincide, $|u|_w = |v|_w$.

**Example**

$A = \{a, b\}$, $u = abababaabb$, $v = aabbababbab$,

- $u \equiv_2 v$ because $\text{pref}_1(u) = a = \text{pref}_1(v)$, $\ldots$, and $|u|_{aa} = 1 = |v|_{aa}, |u|_{ab} = 3 = |v|_{ab}, \ldots$  
- $u \not\equiv_3 v$ because $\text{suf}_2(u) = bb \neq ab = \text{suf}_2(v)$
- $abcababb \equiv_3 ababcabb$
$k$-abelian equivalence

Let $k \geq 1$ be an integer. Two words $u$ and $v$ in $A^+$ are $k$-abelian equivalent, denoted by $u \equiv_k v$, if

- $\text{pref}_{k-1}(u) = \text{pref}_{k-1}(v)$,
- $\text{suf}_{k-1}(u) = \text{suf}_{k-1}(v)$,
- for all $w \in A^k$, the number of occurrences of $w$ in $u$ and in $v$ coincide, $|u|_w = |v|_w$.

Remark

- $\equiv_k$ is an equivalence relation
- $u \equiv_k v \implies u \equiv_{k-1} v$, $\forall k \geq 1$
- $u = v \iff u \equiv_k v$, $\forall k \geq 1$
2-abelian complexity of \( t \)

The first values of the 2-abelian complexity of the Thue-Morse word

\[
P^{(2)}_t(n) = \#\{ \text{factors of length } n \text{ of } t \}/\equiv_2
\]

are

\[
(P^{(2)}_t(n))_{n \geq 0} = (1, 2, 4, 6, 8, 6, 8, 10, 8, 6, 8, 8, 10, 10, 10, 10, 10, 8, 10, 10, 10, 8, 10, 10, 12, 12, 10, 12, 12, 10, 10, 10, 10, 12, 12, 10, 8, 10, 10, 12, 12, 12, 10, 10, 12, 12, 12, 12, 12, 10, 8, 10, 10, 12, 12, 10, 10, 8, 10, 8, 10, 10, 12, 12, 10, 12, 12, 14, 12, 10, 10, 12, 10, 8, 10, 8, 10, 10, 12, 12, 12, 12, 12, 14, 12, 10, 8, 10, 12, 14, 12, 12, 12, 10, 12, 12, 12, 12, 12, 12, 10, 12, \ldots
\]
2-abelian complexity of $t$

The first values of the 2-abelian complexity of the Thue-Morse word

$$P_t^{(2)}(n) = \#\{\text{factors of length } n \text{ of } t\}/\equiv_2$$

are

$$(P_t^{(2)}(n))_{n \geq 0} = (1, 2, 4, 6, 8, 10, 8, 6, 8, 10, 10, 12, 12, 10, 12, 10, 8, 10, 10,$

$10, 8, 8, 6, 8, 10, 10, 8, 10, 12, 12, 10, 12, 12, 10, 8, 10, 10, 12, 12, 10, 12, 12, 10, 12,$

$12, 12, 12, 12, 14, 12, 10, 8, 10, 12, 12, 10, 10, 8, 8, 6, 8, 10,$

$10, 8, 10, 12, 12, 10, 12, 12, 12, 12, 14, 12, 12, 10, 8, 10, 12, 14,$

$12, 14, 16, 14, 12, 14, 14, 14, 12, 12, 12, 12, 10, 12, 12, \ldots$
A sequence \((x_n)_{n \geq 0}\) (over \(\mathbb{Z}\)) is \(k\)-regular of its \(\mathbb{Z}\)-module generated by its \(k\)-kernel

\[
K = \{(x_{k^e n + r})_{n \geq 0} \mid e \geq 0, r < k^e\}
\]

is finitely generated.


**Example**

The 2-kernel of \(t\) is

\[
K = \{(t_{2^e n + r})_{n \geq 0} \mid e \geq 0, r < 2^e\}
\]

\[= \{t, \bar{t}\}\]

where \(\bar{t} = (1 - t_n)_{n \geq 0}\).
A sequence \((x_n)_{n \geq 0}\) (over \(\mathbb{Z}\)) is \(k\)-regular of its \(\mathbb{Z}\)-module generated by its \(k\)-kernel

\[
\mathcal{K} = \{(x_{k^e n + r})_{n \geq 0} \mid e \geq 0, r < k^e\}
\]

is finitely generated.


**Theorem (Eilenberg)**

A sequence \((x_n)_{n \geq 0}\) is \(k\)-automatic iff its \(k\)-kernel is finite.
### Theorem (Madill, Rampersad)

The abelian complexity of the paperfolding word

$$0010011000110110001001110011011\cdots$$

is a 2-regular sequence.

### Proposition (Karhumäki, Saarela, Zamboni)

The abelian complexity of the period doubling word, obtained as the fixed point of $\mu : 0 \mapsto 01, 1 \mapsto 00$, is a 2-regular sequence.

### Question

Is the abelian complexity of a $k$-automatic sequence always $k$-regular?
Conjecture

The 2-abelian complexity of $t$ is 2-regular.

**Notation**: $x_{2^e+r} = (P_t^{(2)}(2^e n + r))_{n \geq 0}$.

We conjecture the following relations (Mathematica experiments):

\[
\begin{align*}
  x_5 &= x_3 \\
  x_9 &= x_3 \\
  x_{12} &= -x_6 + x_7 + x_{11} \\
  x_{13} &= x_7 \\
  x_{16} &= x_8 \\
  x_{17} &= x_3 \\
  x_{18} &= x_{10} \\
  x_{20} &= -x_{10} + x_{11} + x_{19} \\
  x_{21} &= x_{11} \\
  x_{22} &= -x_3 - 2x_6 + x_7 + 3x_{10} + x_{11} - x_{19} \\
  x_{23} &= -x_3 - 3x_6 + 2x_7 + 3x_{10} + x_{11} - x_{19} \\
  x_{24} &= -x_3 + x_7 + x_{10} \\
  x_{25} &= x_7 \\
  x_{26} &= -x_3 + x_7 + x_{10} \\
  x_{27} &= -2x_3 + x_7 + 3x_{10} - x_{19} \\
  x_{28} &= -2x_3 + x_7 + 3x_{10} - x_{14} + x_{15} - x_{19} \\
  x_{29} &= x_{15} \\
  x_{30} &= -x_3 + 3x_6 - x_7 - x_{10} - x_{11} + x_{15} + x_{19} \\
  x_{31} &= -3x_3 + 6x_6 - 2x_{11} - 3x_{14} + 2x_{15} + x_{19}
\end{align*}
\]
We also conjecture the following relations

\[
\begin{align*}
x_{32} &= x_8 \\
x_{33} &= x_3 \\
x_{34} &= x_{10} \\
x_{35} &= x_{11} \\
x_{36} &= -x_{10} + x_{11} + x_{19} \\
x_{37} &= x_{19} \\
x_{38} &= -x_{3} + x_{10} + x_{19} \\
x_{39} &= -x_{3} + x_{11} + x_{19} \\
x_{40} &= -x_{3} + x_{10} + x_{11} \\
x_{41} &= x_{11} \\
x_{42} &= -x_{3} + x_{10} + x_{11} \\
x_{43} &= -2x_{3} + 3x_{10} \\
x_{44} &= -2x_{3} - x_{6} + x_{7} + 3x_{10} \\
x_{45} &= -x_{3} - 3x_{6} + 2x_{7} + 3x_{10} + x_{11} - x_{19} \\
x_{46} &= -2x_{3} - 3x_{6} + 2x_{7} + 5x_{10} + x_{11} - 2x_{19} \\
x_{47} &= -2x_{3} + x_{7} + 3x_{10} - x_{19} \\
x_{48} &= -x_{3} + x_{7} + x_{10} \\
x_{49} &= x_{7} \\
x_{50} &= -x_{3} + x_{7} + x_{10} \\
x_{51} &= -x_{3} - 3x_{6} + 2x_{7} + 3x_{10} + x_{11} - x_{19} \\
x_{52} &= -2x_{3} - 3x_{6} + 2x_{7} + 5x_{10} + x_{11} - 2x_{19} \\
x_{53} &= -2x_{3} + x_{7} + 3x_{10} - x_{19} \\
x_{54} &= -4x_{3} + 3x_{6} + x_{7} + 3x_{10} - x_{11} - 2x_{14} + x_{15} \\
x_{55} &= -4x_{3} + 3x_{6} + x_{7} + 3x_{10} - x_{11} - 3x_{14} + 2x_{15} \\
x_{56} &= -x_{3} + x_{10} + x_{15} \\
x_{57} &= x_{15} \\
x_{58} &= -x_{3} + x_{10} + x_{15} \\
x_{59} &= -2x_{3} + 3x_{6} - x_{7} - x_{11} + x_{15} + x_{19} \\
x_{60} &= -4x_{3} + 6x_{6} + x_{10} - 2x_{11} - 3x_{14} + 2x_{15} + x_{19} \\
x_{61} &= -3x_{3} + 6x_{6} - 2x_{11} - 3x_{14} + 2x_{15} + x_{19} \\
x_{62} &= -x_{3} + 3x_{6} - x_{7} - x_{10} - x_{11} + x_{15} + x_{19} \\
x_{63} &= x_{15}
\end{align*}
\]

If the conjecture is true, then any sequence \( x_n \) for \( n \geq 32 \) is a linear combination of \( x_1, x_2, \ldots, x_{19} \).
Proposition

For all $n \geq 0$, $P_t^{(2)}(2n + 1) = P_t^{(2)}(4n + 1)$.

\[
\begin{align*}
    x_{32} &= x_8 \\
    x_{33} &= x_3 \\
    x_{34} &= x_{10} \\
    x_{35} &= x_{11} \\
    x_{36} &= -x_{10} + x_{11} + x_{19} \\
    x_{37} &= x_{19} \\
    x_{38} &= -x_3 + x_{10} + x_{19} \\
    x_{39} &= -x_3 + x_{11} + x_{19} \\
    x_{40} &= -x_3 + x_{10} + x_{11} \\
    x_{41} &= x_{11} \\
    x_{42} &= -x_3 + x_{10} + x_{11} \\
    x_{43} &= -2x_3 + 3x_{10} \\
    x_{44} &= -2x_3 - x_6 + x_7 + 3x_{10} \\
    x_{45} &= x_{23} \\
    x_{46} &= -2x_3 - 3x_6 + 2x_7 + 5x_{10} + x_{11} - 2x_{19} \\
    x_{47} &= -2x_3 + x_7 + 3x_{10} - x_{19} \\
    x_{48} &= -x_3 + x_7 + x_{10} \\
    x_{49} &= x_7 \\
    x_{50} &= -x_3 + x_7 + x_{10} \\
    x_{51} &= -x_3 - 3x_6 + 2x_7 + 3x_{10} + x_{11} - x_{19} \\
    x_{52} &= -2x_3 - 3x_6 + 2x_7 + 5x_{10} + x_{11} - 2x_{19} \\
    x_{53} &= x_{27} \\
    x_{54} &= -4x_3 + 3x_6 + x_7 + 3x_{10} - x_{11} - 2x_{14} + x_{15} \\
    x_{55} &= -4x_3 + 3x_6 + x_7 + 3x_{10} - x_{11} - 3x_{14} + 2x_{15} \\
    x_{56} &= -x_3 + x_{10} + x_{15} \\
    x_{57} &= x_{15} \\
    x_{58} &= -x_3 + x_{10} + x_{15} \\
    x_{59} &= -2x_3 + 3x_6 - x_7 - x_{11} + x_{15} + x_{19} \\
    x_{60} &= -4x_3 + 6x_6 + x_{10} - 2x_{11} - 3x_{14} + 2x_{15} + x_{19} \\
    x_{61} &= x_{31} \\
    x_{62} &= -x_3 + 3x_6 - x_7 - x_{10} - x_{11} + x_{15} + x_{19} \\
    x_{63} &= x_{15}
\end{align*}
\]
Another approach

Consider the function

$$f : \mathbb{N} \to \mathbb{N}^4, \ n \mapsto \begin{pmatrix} p_n \mid 00 \\ p_n \mid 01 \\ p_n \mid 10 \\ p_n \mid 11 \end{pmatrix}$$

where $p_n$ is the prefix of length $n$ of the Thue-Morse word.

Properties

- $f(3 \cdot 2^i + 1) = (2^{i-1}, 2^i, 2^i, 2^{i-1})$
- $f(3 \cdot 2^i) = \begin{cases} (2^{i-1} - 1, 2^i, 2^i, 2^{i-1}) & \text{if } i \text{ is odd} \\ (2^{i-1}, 2^i, 2^i - 1, 2^{i-1}) & \text{if } i \text{ is even} \end{cases}$
Property

The function $f_{01}: n \mapsto |p_n|_{01}$ is 2-regular.

<table>
<thead>
<tr>
<th>$t$</th>
<th>0 1 1 0 1 0 0 1 1 0 0 1 0 1 1 1 ...</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(a_n)$</td>
<td>1 0 0 1 0 0 1 0 0 0 1 0 1 0 0 ...</td>
</tr>
<tr>
<td>$(b_n)$</td>
<td>1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 ...</td>
</tr>
<tr>
<td>$(f_{01}(n))$</td>
<td>1 1 1 2 2 2 3 3 3 3 4 4 5 5 5 ...</td>
</tr>
</tbody>
</table>

Remark

The convolution of two $k$-regular sequences $(a_n)_{n \geq 0}$ and $(b_n)_{n \geq 0}$

$$(a_n)_{n \geq 0} \ast (b_n)_{n \geq 0} = \left( \sum_{i+j=n} a(i)b(j) \right)_{n \geq 0}$$

is a $k$-regular sequence.
Can we find a nice and useful property of the function $f_{01}$? For example, is the sequence $(f_{01}(n))$ 2-synchronized? 

\[ \{ (\text{rep}_2(n), \text{rep}_2(f_{01}(n))) : n \in \mathbb{N} \} \] is accepted by a DFA?
Why such a property would be useful?

If \((f_{01}(n))\) is 2-synchronized,

1. \(\{(\text{rep}_2(n), \text{rep}_2(f_{01}(n))) : n \in \mathbb{N}\}\) is accepted by a DFA.
2. \(L = \{(\text{rep}_2(\ell), \text{rep}_2(f_{01}(n + \ell) - f_{01}(n)) : \ell, n \in \mathbb{N}\}\) is accepted by a DFA.
3. \(\ell \mapsto \#\{(\text{rep}_2(\ell), -) \in L\}\) forms a 2-regular sequence.

Theorem (Charlier, Rampersad, Shallit)

Let \(A, B \subset \mathbb{N}\). If the language

\[\{(\text{rep}_k(n), \text{rep}_k(m)) : (n, m) \in A \times B\}\]

is accepted by a DFA, then \(n \mapsto \#\{(\text{rep}_k(n), -) \in L\}\) forms a \(k\)-regular sequence.
$(f_{01}(n))$ is not 2-synchronized

- Assume $(f_{01}(n))$ is 2-synchronized.
- Then $(f_{01}(n) - \frac{n}{3})$ is 2-synchronized.
- For $n$ with $\text{rep}_2(n) = (10)^{4\ell}$, $f_{01}(n) - \frac{n}{3} = -\frac{2\ell}{3}$.
- For such $n$, the subsequence has logarithmic growth and is 2-synchronized.
- Any non-increasing $k$-synchronized sequence is either constant or linear.
- So $(f_{01}(n))$ is not 2-synchronized.
$k$-regular seq.

$k$-synchronized seq.

$k$-automatic sequences

= bounded & $k$-synchronized seq.