2-ABELIAN COMPLEXITY OF THE THUE-MORSE SEQUENCE (WORK IN PROGRESS)

Michel Rigo, joint work with E. Vandomme

http://www.discmath.ulg.ac.be/
http://orbi.ulg.ac.be/

Université de Liège

Basics

The *Thue–Morse* word $\mathbf{t}$ is the infinite word $\lim_{n\to\infty} f^n(a)$ where

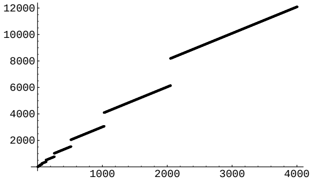
$$f : a \mapsto ab, \quad b \mapsto ba$$

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

The factor complexity of the Thue–Morse word is well-known

$$p_{\mathbf{t}}(0) = 1, \quad p_{\mathbf{t}}(1) = 2, \quad p_{\mathbf{t}}(2) = 4,$$

$$p_{\mathbf{t}}(n) = \begin{cases} 4n - 2\cdot 2^m - 4 & \text{if } 2\cdot 2^m < n \leq 3\cdot 2^m \\ 2n + 4\cdot 2^m - 2 & \text{if } 3\cdot 2^m < n \leq 4\cdot 2^m \end{cases}$$



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

Basics

Let $k \geq 1$ be an integer. Two words u and v in A^+ are k-*abelian equivalent*, in symbols $u \equiv_{a,k} v$, if

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

Remark

$\equiv_{a,k}$ is an equivalence relation

$$u = v \Rightarrow u \equiv_{a,k} v \Rightarrow u \equiv_a v$$

$$u = v \Leftrightarrow u \equiv_{a,k} v, \forall k \geq 1.$$

Basics

- J. Karhumäki, Generalized Parikh Mappings and Homomorphisms, Information and control **47**, 155–165 (1980).
- M. Huova, J. Karhumäki, A. Saarela, K. Saari, Local squares, periodicity and finite automata, Rainbow of Computer Science, 90–101, Springer, (2011).
- M. Huova, J. Karhumäki, A. Saarela, Problems in between words and abelian words: k-abelian avoidability, Special issue of TCS.

A few examples

$$abbabaabb \equiv_{a,2} aabbabbab$$

$$|w|_{aa} = 1, \quad |w|_{ab} = 3, \quad |w|_{ba} = 2, \quad |w|_{bb} = 2$$

but the two words are not 3-abelian equivalent,
$|u|_{aba} = 1$ and $|v|_{aba} = 0$.
Note that

$$\sum_{f \in A^k} |w|_f = |w| - k + 1$$

$$abcababb \equiv_{a,3} ababcabb$$

Number of equivalence classes for 2-abelian factors of length n occurring in the Thue–Morse word,

$$a_{2,\mathbf{t}}(n) := \mathcal{P}_{\mathbf{t}}^{(2)}(n) = \#(\mathcal{F}_{\mathbf{t}}(n)/ \equiv_{a,k})$$

$$\begin{aligned}(a_{2,\mathbf{t}}(n))_{n\geq 0} =& 1, 2, 4, 6, 8, 6, 8, 10, 8, 6, 8, 8, 10, 10,\\ & 10, 8, 8, 6, 8, 10, 10, 8, 10, 12, 12, 10, 12, 12, 10, 8, 10, 10,\\ & 8, 6, 8, 8, 10, 10, 12, 12, 10, 8, 10, 12, 14, 12, 12, 12, 12, 10,\\ & 12, 12, 12, 12, 14, 12, 10, 8, 10, 12, 12, 10, 10, 8, 8, 6, 8, 10,\\ & 10, 8, 10, 12, 12, 10, 12, 12, 12, 12, 14, 12, 10, 8, 10, 12, 14,\\ & 12, 14, 16, 14, 12, 14, 14, 14, 12, 12, 12, 12, 10, 12, 12, \ldots\end{aligned}$$

Questions

- Is the sequence $(a_{2,\mathbf{t}}(n))_{n\geq 0}$ bounded?
- How to compute (easily) these values?
- Is there a structure behind?
- What is general about any q-automatic sequence?

The 6 factors of length 3

$$aab,\ aba,\ abb,\ baa,\ bab,\ bba$$

occur in the Thue–Morse word (aaa and bbb do not occur) and are pairwise 2-abelian non-equivalent.
We consider vectors of $\mathbb{N}^{10}$ for any word $u = u_1 u_2 \cdots u_{\ell-1} u_\ell$

$$\Psi(u) = \begin{pmatrix} |u_1|_a \\ |u_1|_b \\ |u|_{aa} \\ |u|_{ab} \\ |u|_{ba} \\ |u|_{bb} \\ |u_{\ell-1} u_\ell|_{aa} \\ |u_{\ell-1} u_\ell|_{ab} \\ |u_{\ell-1} u_\ell|_{ba} \\ |u_{\ell-1} u_\ell|_{bb} \end{pmatrix}$$

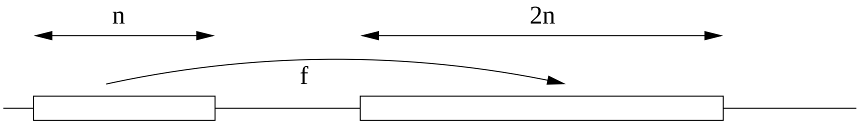
almost similar to $\Pi_2 : \{a, b\}^* \to \mathbb{N}^8$ introduced in Juhani's talk.

$u \equiv_{a,2} v$ if and only if $\Psi(u) \sim \Psi(v)$, i.e.,

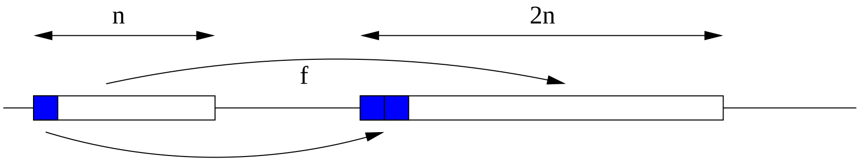
- the first six components of $\Psi(u)$ and $\Psi(v)$ coïncide
- $[\Psi(u)]_7 + [\Psi(u)]_9 = [\Psi(v)]_7 + [\Psi(v)]_9$
- $[\Psi(u)]_8 + [\Psi(u)]_{10} = [\Psi(v)]_8 + [\Psi(v)]_{10}$

	aab	aba	abb	baa	bab	bba
$\lvert u_1 \rvert_a$	1	1	1	0	0	0
$\lvert u_1 \rvert_b$	0	0	0	1	1	1
$\lvert u \rvert_{aa}$	1	0	0	1	0	0
$\lvert u \rvert_{ab}$	1	1	1	0	1	0
$\lvert u \rvert_{ba}$	0	1	0	1	1	1
$\lvert u \rvert_{bb}$	0	0	1	0	0	1
$\lvert u_{\ell-1} u_\ell \rvert_{aa}$	0	0	0	1	0	0
$\lvert u_{\ell-1} u_\ell \rvert_{ab}$	1	0	0	0	1	0
$\lvert u_{\ell-1} u_\ell \rvert_{ba}$	0	1	0	0	0	1
$\lvert u_{\ell-1} u_\ell \rvert_{bb}$	0	0	1	0	0	0

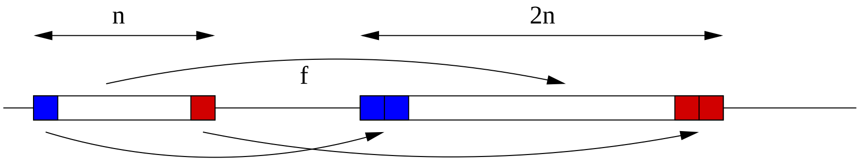
Computation of $a_{2,\mathbf{t}}(\ell)$, first for ℓ odd, i.e., $\ell = 2n - 1$



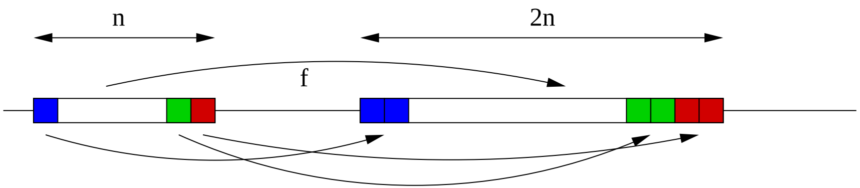
Computation of $a_{2,\mathbf{t}}(\ell)$, first for ℓ odd, i.e., $\ell = 2n - 1$



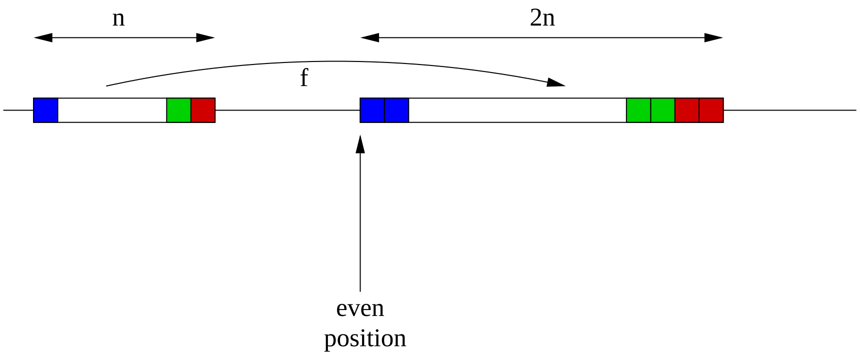
Computation of $a_{2,\mathbf{t}}(\ell)$, first for ℓ odd, i.e., $\ell = 2n - 1$



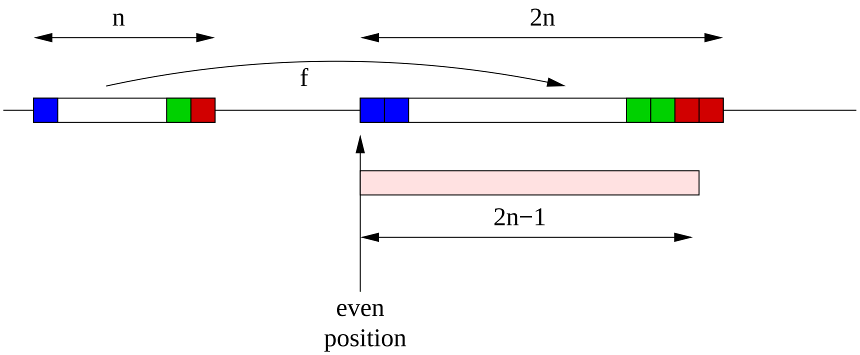
Computation of $a_{2,\mathbf{t}}(\ell)$, first for ℓ odd, i.e., $\ell = 2n - 1$



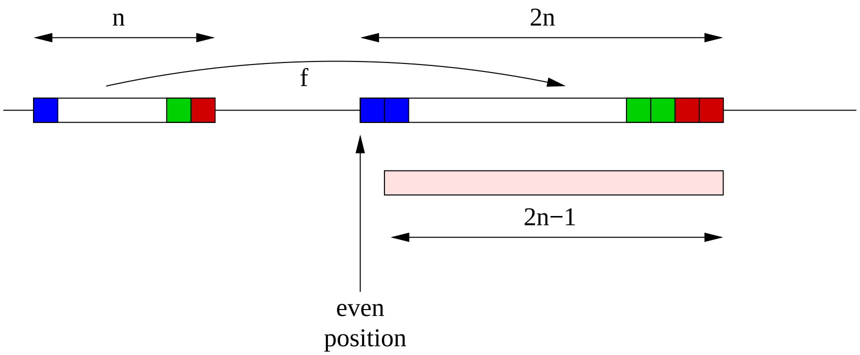
Computation of $a_{2,\mathbf{t}}(\ell)$, first for ℓ odd, i.e., $\ell = 2n - 1$



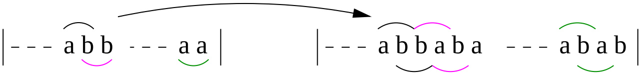
Computation of $a_{2,\mathbf{t}}(\ell)$, first for ℓ odd, i.e., $\ell = 2n - 1$



Computation of $a_{2,\mathbf{t}}(\ell)$, first for ℓ odd, i.e., $\ell = 2n - 1$



We know precisely what's happening



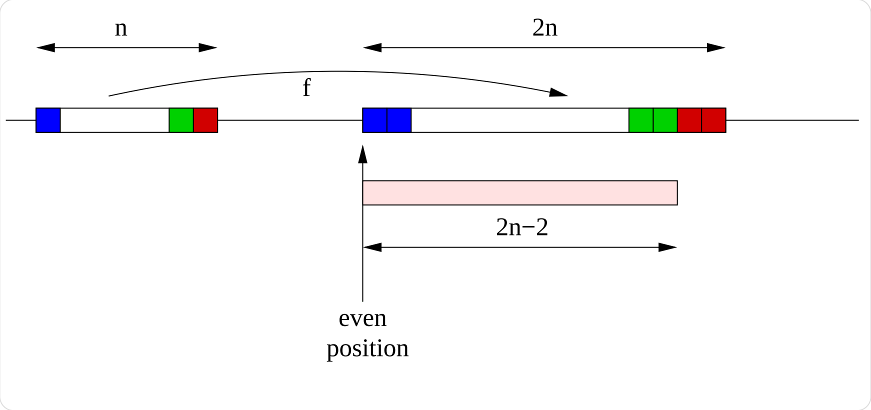
Given a vector corresponding to a factor of length n occurring in $\mathbf{t}$, these two matrices produce vectors corresponding to factors of length $2n-1$ occurring respectively in an *even* and *odd* position

$$mM = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 & -1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 & -1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

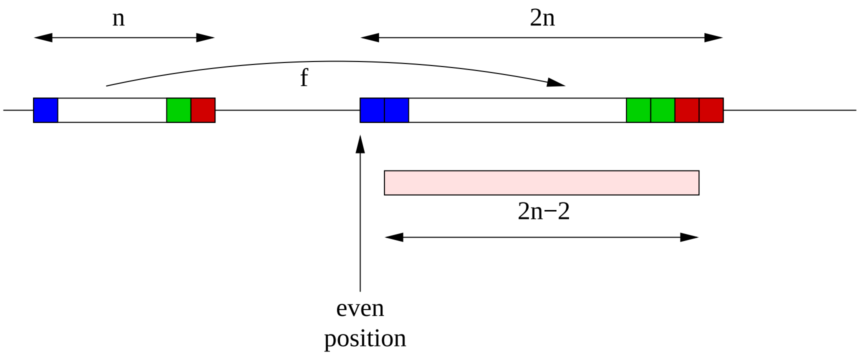
$$mN = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 &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

each block of length 2 produces roughly 2 blocks of length 2.

Computation of $a_{2,\mathbf{t}}(\ell)$, now for ℓ even, i.e., $\ell = 2n - 2$



Computation of $a_{2,\mathbf{t}}(\ell)$, now for ℓ even, i.e., $\ell = 2n - 2$



Given a vector corresponding to a factor of length n occurring in $\mathbf{t}$, these two matrices produce vectors corresponding to factors of length $2n-2$ occurring respectively in an *even* and *odd* position

$$mC = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 & -1 & 0 & -1 & -1 \\ 0 & 1 & 1 & 1 & 0 & 1 & -1 & -1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$mD = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & -1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 & -1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

Lemma (from n to $2n-1$)

Let $\mathbf{y}, \mathbf{z} \in \mathbb{N}^{10}$.

$$\mathbf{y} \sim \mathbf{z} \Rightarrow \begin{cases} mM\mathbf{y} \sim mM\mathbf{z} \\ mN\mathbf{y} \sim mN\mathbf{z}. \end{cases}$$

he converse does not hold in general: $abaab \not\equiv_{a,2} ababb$ but $abbaababb(a) \equiv_{a,2} abbaabbab(a)$ and $(a)bbaababba \equiv_{a,2} (a)bbaabbaba$.

Lemma (from n to $2n-2$)

Let $\mathbf{y}, \mathbf{z} \in \mathbb{N}^{10}$.

$$\mathbf{y} \sim \mathbf{z} \Rightarrow mD\mathbf{y} \sim mD\mathbf{z}$$

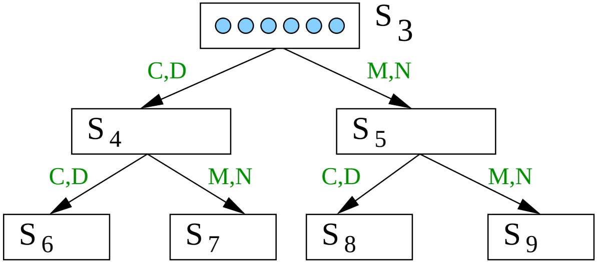
If $\mathbf{y} \sim \mathbf{z}$ and $\mathbf{y} \neq \mathbf{z}$, this means that the corresponding second to last letters are different, hence

$$(\mathbf{y} \sim \mathbf{z} \wedge \mathbf{y} \neq \mathbf{z}) \Rightarrow mC\mathbf{y} \not\sim mC\mathbf{z}.$$

$$S_3 = \{\mathbf{v} \in \mathbb{N}^{10} \mid \exists u \in A^3 : \mathbf{v} = \Psi(u) \wedge u \text{ occurs in } \mathbf{t}\}$$

$$S_4 = \{mC\mathbf{v},\ mD\mathbf{v} \mid \mathbf{v} \in S_3\}$$

$$S_5 = \{mM\mathbf{v},\ mN\mathbf{v} \mid \mathbf{v} \in S_3\}$$

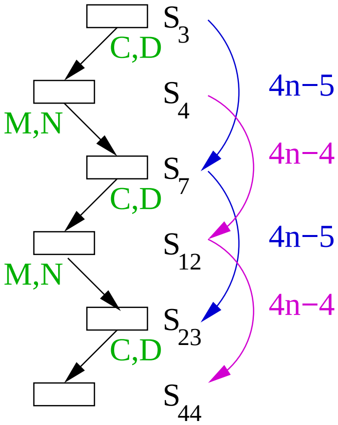


$$a_{2,\mathbf{t}}(n) = \#(S_n/\sim).$$

Conjecture

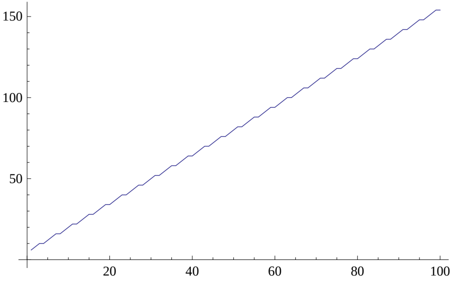
The sequence $((a_{2,\mathbf{t}}(n))_{n\geq 0}$ is unbounded.

Consider the subsequence corresponding to the following picture



We get the following

n	$a_{2,\mathbf{t}}(n)$
3	6
4	8
7	10
12	10
23	12
44	14
87	16
172	16
343	18
684	20
1367	22
2732	22
5463	24
10924	26
21847	28
43692	28
87383	30
174764	32
349527	34
699052	34
1398103	36
2796204	38
5592407	40
11184812	40
22369623	42
44739244	44
89478487	46
178956972	46
357913943	48
715827884	50
1431655767	52
2863311532	52



6, 8, 10, 10, 12, 14, 16, 16, 18, 20, 22, 22, 24, 26, 28, 28, 30, 32, 34, 34, 36, 38, 40, 40, 42, 44, 46, 46, 48, 50, 52, 52, 54, 56, 58, 58, 60, 62, 64, 64, 66, 68, 70, 70, 72, 74, 76, 76, 78, 80, 82, 82, 84, 86, 88, 88, 90, 92, 94, 94, 96, 98, 100, 100, 102, 104, 106, 106, 108, 110, 112, 112, 114, 116, 118, 118, 120, 122, 124, 124, 126, 128, 130,

The sequence seems to satisfy the relation

$$y_{n+5} = y_{n+4} + y_{n+1} - y_n$$

and

$$y_n = \frac{3}{2}n + \frac{25 + (-1)^n - (1-i)(-i)^n - (1+i)i^n}{4}.$$

A sequence $(x_n)_{n\geq 0}$ (over $\mathbb{Z}$) is k-*regular* if the $\mathbb{Z}$-module generated by its k-kernel

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

is finitely generated.

J.-P. Allouche, J. Shallit, The ring of k-regular sequences, *Theoret. Comput. Sci.* **98** (1992)

Proposition (Eilenberg)

A sequence $(x_n)_{n\geq 0}$ is k-automatic if and only if its k-kernel is finite.

Notation:

$$\boxed{\mathbf{x}_{2^e+r} = (a_{2,\mathbf{t}}(2^e n + r))_{n\geq 0}.}$$

Section 6 Recognizing a k-regular sequence in the paper:

J.-P. Allouche, J. Shallit, The ring of k-regular sequences. II, *Theoret. Comput. Sci.* **307** (2003).

- We compute the first $N = 100$ terms of the first 63 sequences in the 2-kernel of the sequence $\mathbf{a} = (a_{2,\mathbf{t}}(n))_{n\geq 0} \in \mathbb{Z}^{\mathbb{N}}$,

  $$\{\mathbf{a}, \mathbf{x}_2 = (\mathbf{t}_{2n}), \mathbf{x}_3 = (\mathbf{t}_{2n+1}), \mathbf{x}_4 = (\mathbf{t}_{4n}), \ldots, \mathbf{x}_{63} = (\mathbf{t}_{32n+31})\}$$

- $j = 1$.
- Select the first sequence $\mathbf{a}$.
- At step j, $r < j$ sequences have been selected, take sequence $\mathbf{x}_j$ from $\mathcal{K}$, check on the first N elements whether $\mathbf{x}_j$ *seems to be* a combination of the selected ones. If not, select this new sequence.
- $j \leftarrow j + 1$, until $j = 63$.

Sequences $\mathbf{x}_{32}, \ldots, \mathbf{x}_{63}$ are all combinations of $\mathbf{a}, \mathbf{x}_2, \ldots, \mathbf{x}_{31}$ (checked for some $N > 10000$).

We conjecture the following relations (`Mathematica` experiments)

$$
\begin{aligned}
\mathbf{x}_5 &= \mathbf{x}_3\\
\mathbf{x}_9 &= \mathbf{x}_3\\
\mathbf{x}_{12} &= -\mathbf{x}_6 + \mathbf{x}_7 + \mathbf{x}_{11}\\
\mathbf{x}_{13} &= \mathbf{x}_7\\
\mathbf{x}_{16} &= \mathbf{x}_8\\
\mathbf{x}_{17} &= \mathbf{x}_3\\
\mathbf{x}_{18} &= \mathbf{x}_{10}\\
\mathbf{x}_{20} &= -\mathbf{x}_{10} + \mathbf{x}_{11} + \mathbf{x}_{19}\\
\mathbf{x}_{21} &= \mathbf{x}_{11}\\
\mathbf{x}_{22} &= -\mathbf{x}_3 - 2\mathbf{x}_6 + \mathbf{x}_7 + 3\mathbf{x}_{10} + \mathbf{x}_{11} - \mathbf{x}_{19}\\
\mathbf{x}_{23} &= -\mathbf{x}_3 - 3\mathbf{x}_6 + 2\mathbf{x}_7 + 3\mathbf{x}_{10} + \mathbf{x}_{11} - \mathbf{x}_{19}\\
\mathbf{x}_{24} &= -\mathbf{x}_3 + \mathbf{x}_7 + \mathbf{x}_{10}\\
\mathbf{x}_{25} &= \mathbf{x}_7\\
\mathbf{x}_{26} &= -\mathbf{x}_3 + \mathbf{x}_7 + \mathbf{x}_{10}\\
\mathbf{x}_{27} &= -2\mathbf{x}_3 + \mathbf{x}_7 + 3\mathbf{x}_{10} - \mathbf{x}_{19}\\
\mathbf{x}_{28} &= -2\mathbf{x}_3 + \mathbf{x}_7 + 3\mathbf{x}_{10} - \mathbf{x}_{14} + \mathbf{x}_{15} - \mathbf{x}_{19}\\
\mathbf{x}_{29} &= \mathbf{x}_{15}\\
\mathbf{x}_{30} &= -\mathbf{x}_3 + 3\mathbf{x}_6 - \mathbf{x}_7 - \mathbf{x}_{10} - \mathbf{x}_{11} + \mathbf{x}_{15} + \mathbf{x}_{19}\\
\mathbf{x}_{31} &= -3\mathbf{x}_3 + 6\mathbf{x}_6 - 2\mathbf{x}_{11} - 3\mathbf{x}_{14} + 2\mathbf{x}_{15} + \mathbf{x}_{19}
\end{aligned}
$$

We also conjecture the following relations

$$
\begin{aligned}
\mathbf{x}_{32} &= \mathbf{x}_8\\
\mathbf{x}_{33} &= \mathbf{x}_3\\
\mathbf{x}_{34} &= \mathbf{x}_{10}\\
\mathbf{x}_{35} &= \mathbf{x}_{11}\\
\mathbf{x}_{36} &= -\mathbf{x}_{10}+\mathbf{x}_{11}+\mathbf{x}_{19}\\
\mathbf{x}_{37} &= \mathbf{x}_{19}\\
\mathbf{x}_{38} &= -\mathbf{x}_3+\mathbf{x}_{10}+\mathbf{x}_{19}\\
\mathbf{x}_{39} &= -\mathbf{x}_3+\mathbf{x}_{11}+\mathbf{x}_{19}\\
\mathbf{x}_{40} &= -\mathbf{x}_3+\mathbf{x}_{10}+\mathbf{x}_{11}\\
\mathbf{x}_{41} &= \mathbf{x}_{11}\\
\mathbf{x}_{42} &= -\mathbf{x}_3+\mathbf{x}_{10}+\mathbf{x}_{11}\\
\mathbf{x}_{43} &= -2\mathbf{x}_3+3\mathbf{x}_{10}\\
\mathbf{x}_{44} &= -2\mathbf{x}_3-\mathbf{x}_6+\mathbf{x}_7+3\mathbf{x}_{10}\\
\mathbf{x}_{45} &= -\mathbf{x}_3-3\mathbf{x}_6+2\mathbf{x}_7+3\mathbf{x}_{10}+\mathbf{x}_{11}-\mathbf{x}_{19}\\
\mathbf{x}_{46} &= -2\mathbf{x}_3-3\mathbf{x}_6+2\mathbf{x}_7+5\mathbf{x}_{10}+\mathbf{x}_{11}-2\mathbf{x}_{19}\\
\mathbf{x}_{47} &= -2\mathbf{x}_3+\mathbf{x}_7+3\mathbf{x}_{10}-\mathbf{x}_{19}\\
\mathbf{x}_{48} &= -\mathbf{x}_3+\mathbf{x}_7+\mathbf{x}_{10}\\
\mathbf{x}_{49} &= \mathbf{x}_7\\
\mathbf{x}_{50} &= -\mathbf{x}_3+\mathbf{x}_7+\mathbf{x}_{10}\\
\mathbf{x}_{51} &= -\mathbf{x}_3-3\mathbf{x}_6+2\mathbf{x}_7+3\mathbf{x}_{10}+\mathbf{x}_{11}-\mathbf{x}_{19}\\
\mathbf{x}_{52} &= -2\mathbf{x}_3-3\mathbf{x}_6+2\mathbf{x}_7+5\mathbf{x}_{10}+\mathbf{x}_{11}-2\mathbf{x}_{19}\\
\mathbf{x}_{53} &= -2\mathbf{x}_3+\mathbf{x}_7+3\mathbf{x}_{10}-\mathbf{x}_{19}\\
\mathbf{x}_{54} &= -4\mathbf{x}_3+3\mathbf{x}_6+\mathbf{x}_7+3\mathbf{x}_{10}-\mathbf{x}_{11}-2\mathbf{x}_{14}+\mathbf{x}_{15}\\
\mathbf{x}_{55} &= -4\mathbf{x}_3+3\mathbf{x}_6+\mathbf{x}_7+3\mathbf{x}_{10}-\mathbf{x}_{11}-3\mathbf{x}_{14}+2\mathbf{x}_{15}\\
\mathbf{x}_{56} &= -\mathbf{x}_3+\mathbf{x}_{10}+\mathbf{x}_{15}\\
\mathbf{x}_{57} &= \mathbf{x}_{15}\\
\mathbf{x}_{58} &= -\mathbf{x}_3+\mathbf{x}_{10}+\mathbf{x}_{15}\\
\mathbf{x}_{59} &= -2\mathbf{x}_3+3\mathbf{x}_6-\mathbf{x}_7-\mathbf{x}_{11}+\mathbf{x}_{15}+\mathbf{x}_{19}\\
\mathbf{x}_{60} &= -4\mathbf{x}_3+6\mathbf{x}_6+\mathbf{x}_{10}-2\mathbf{x}_{11}-3\mathbf{x}_{14}+2\mathbf{x}_{15}+\mathbf{x}_{19}\\
\mathbf{x}_{61} &= -3\mathbf{x}_3+6\mathbf{x}_6-2\mathbf{x}_{11}-3\mathbf{x}_{14}+2\mathbf{x}_{15}+\mathbf{x}_{19}\\
\mathbf{x}_{62} &= -\mathbf{x}_3+3\mathbf{x}_6-\mathbf{x}_7-\mathbf{x}_{10}-\mathbf{x}_{11}+\mathbf{x}_{15}+\mathbf{x}_{19}\\
\mathbf{x}_{63} &= \mathbf{x}_{15}
\end{aligned}
$$

If the conjecture holds, then any $\mathbf{x}_n$ for $n \geq 32$ is a linear combination of $\mathbf{a}, \mathbf{x}_2, \ldots, \mathbf{x}_{19}$.

Conjecture

The sequence $(a_{2,\mathbf{t}}(n))_{n\geq 0}$ is 2-regular.

Example

To get $\mathbf{x}_{75} = (a_{2,\mathbf{t}}(64n+11))_{n\geq 0}$, take every second element in

$$\begin{aligned}(a_{2,\mathbf{t}}(32n+11))_{n\geq 0} &= \mathbf{x}_{43} \\ &= -2\mathbf{x}_3 + 3\mathbf{x}_{10} \\ &= -2(a_{2,\mathbf{t}}(2n+1))_{n\geq 0} + 3(a_{2,\mathbf{t}}(8n+2))_{n\geq 0}.\end{aligned}$$

Hence

$$\begin{aligned}\mathbf{x}_{75} &= (a_{2,\mathbf{t}}(64n+11))_{n\geq 0} \\ &= -2(a_{2,\mathbf{t}}(4n+1))_{n\geq 0} + 3(a_{2,\mathbf{t}}(16n+2))_{n\geq 0} \\ &= -2\mathbf{x}_5 + 3\mathbf{x}_{18}.\end{aligned}$$

Lemma

Let $\mathbf{y}, \mathbf{z} \in \mathbb{N}^{10}$. We have

$$mM\mathbf{y} \sim mM\mathbf{z} \Leftrightarrow mN\mathbf{y} \sim mN\mathbf{z},$$

$$mM\mathbf{y} \sim mN\mathbf{z} \Leftrightarrow mN\mathbf{y} \sim mM\mathbf{z},$$

from factors of length $n+1$ to factors of length $2n+1$.

E.g., $\mathbf{y} = \Psi(aba)$, $\mathbf{z} = \Psi(bab)$, $mM\mathbf{y} \sim mN\mathbf{z}$ and $mN\mathbf{y} \sim mM\mathbf{z}$
$abbaa(b) \equiv_{a,2} (b)aabba$, $(a)bbaab \equiv_{a,2} baabb(a)$.

Lemma

Let $\mathbf{y}, \mathbf{z} \in \mathbb{N}^{10}$. We have

$$mM\mathbf{y} \sim mM\mathbf{z} \Leftrightarrow mN\mathbf{y} \sim mN\mathbf{z},$$

$$mM\mathbf{y} \sim mN\mathbf{z} \Leftrightarrow mN\mathbf{y} \sim mM\mathbf{z},$$

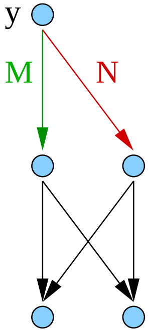
from factors of length $n+1$ to factors of length $2n+1$.

E.g., $\mathbf{y} = \Psi(aba)$, $\mathbf{z} = \Psi(bab)$, $mM\mathbf{y} \sim mN\mathbf{z}$ and $mN\mathbf{y} \sim mM\mathbf{z}$
$abbaa(b) \equiv_{a,2} (b)aabba$, $(a)bbaab \equiv_{a,2} baabb(a)$.

Lemma

Let $\mathbf{y} \in \mathbb{N}^{10}$.

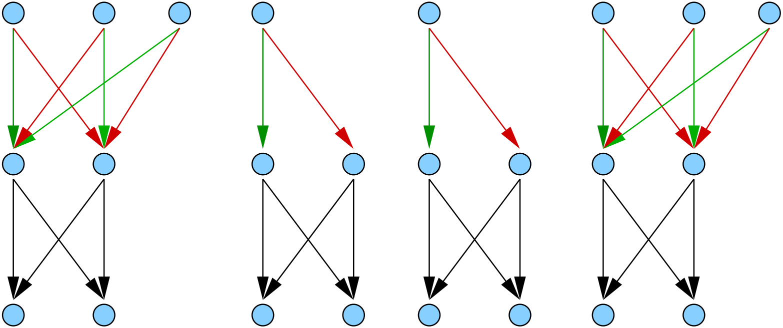
- $mM\mathbf{y} \not\sim mN\mathbf{y}$
- $\#\{mMmM\mathbf{y}, mMmN\mathbf{y}, mNmM\mathbf{y}, mNmN\mathbf{y}\}/\sim = 2.$



from factors of length $n+1$ to factors of length $2n+1$ and $4n+1$.

Proposition

For all n, $a_{2,\mathbf{t}}(2n+1) = a_{2,\mathbf{t}}(4n+1)$.



from factors of length $n+1$ to factors of length $2n+1$ and $4n+1$.

$2n+1$	3	5	7	9	11	13	15	17	19	21	23	25	27	29	31
$4n+1$	5	9	13	17	21	25	29	33	37	41	45	49	53	57	61

Corollary:

$$\boxed{\begin{array}{l} \mathbf{x}_3 = \mathbf{x}_5 = \mathbf{x}_9 = \mathbf{x}_{17} = \mathbf{x}_{33} \\ \mathbf{x}_7 = \mathbf{x}_{13} = \mathbf{x}_{25} = \mathbf{x}_{49} \\ \mathbf{x}_{11} = \mathbf{x}_{21} = \mathbf{x}_{41} \\ \mathbf{x}_{15} = \mathbf{x}_{29} = \mathbf{x}_{57} \\ \mathbf{x}_{19} = \mathbf{x}_{37} \end{array}}$$

We also obtain "new" relations

$$\boxed{\mathbf{x}_{23} = \mathbf{x}_{45}, \quad \mathbf{x}_{27} = \mathbf{x}_{53}, \quad \mathbf{x}_{31} = \mathbf{x}_{61}.}$$

Related work: B. Madill, N. Rampersad, The abelian complexity of the paperfolding word, `arXiv:1208.2856`

$$0010011000110110001001110011011\cdots$$

Theorem

The abelian complexity function of the ordinary paperfolding word is a 2-regular sequence.

Question

Is the abelian complexity function of a q-automatic sequence always q-regular?

We can generalize the question. *Is the k-abelian complexity function of a q-automatic sequence always q-regular?*

Another approach

- V. Bruyère, G. Hansel, C. Michaux, R. Villemaire, Logic and p-recognizable sets of integers, *Bull. Belg. Math. Soc.* **1** (1994).
- J.-P. Allouche, N. Rampersad, J. Shallit, Periodicity, repetitions, and orbits of an automatic sequence, *Theoret. Comput. Sci.* **410** (2009).
- E. Charlier, N. Rampersad, J. Shallit, Enumeration and Decidable Properties of Automatic Sequences, `arXiv:1102.3698`.
- D. Henshall, J. Shallit, Automatic Theorem-Proving in Combinatorics on Words, `arXiv:1203.3758`.
- D. Goc, H. Mousavi, J. Shallit, On the Number of Unbordered Factors, `arXiv:1211.1301`.

We take verbatim *Büchi's theorem* as stated by Charlier, Rampersad and Shallit expressing that k-*automatic sequences are exactly the sequences definable in the first order structure* $\langle \mathbb{N}, +, V_k \rangle$.

Theorem

If we can express a property of a k-automatic sequence $\mathbf{x}$ using quantifiers, logical operations, integer variables, the operations of addition, subtraction, indexing into $\mathbf{x}$, and comparison of integers or elements of $\mathbf{x}$, then this property is decidable.

Let $\mathbf{x}$ be a k-automatic sequence.

- Same factor of length n occurring in position i and j

$$F_{\mathbf{x}}(n,i,j) \equiv (\forall k < n)(\mathbf{x}(i+k) = \mathbf{x}(j+k))$$

- First occurrence of a factor of length n occurring in position i

$$P_{\mathbf{x}}(n,i) \equiv (\forall j < i) \neg F_{\mathbf{x}}(n,i,j)$$

The set $\{(n,i) \mid P_{\mathbf{x}}(n,i) \text{ true}\}$ is k-recognizable and

$$\forall n \geq 0, \quad \#\{i \mid P_{\mathbf{x}}(n,i) \text{ true}\} = p_{\mathbf{x}}(n).$$

Let $\mathbf{x}$ a k-automatic sequence.

- Two factors of length n occurring in position i and j are abelian equivalent

$$A_{\mathbf{x}}(n,i,j) \equiv (\exists \nu \in S_n)(\forall k < n)(\mathbf{x}(i+k) = \mathbf{x}(\nu(j+k)))$$

  The length of the formula is $\simeq n!$ and grows with n.

- First occurrence (up to abelian equivalence) of a factor of length n occurring in position i

$$AP_{\mathbf{x}}(n,i) \equiv (\forall j < i) \neg A_{\mathbf{x}}(n,i,j)$$

For a constant n. The set $\{i \mid AP_{\mathbf{x}}(n,i) \text{ true}\}$ is k-recognizable and

$$\#\{i \mid AP_{\mathbf{x}}(n,i) \text{ true}\} = a_{\mathbf{x}}(n).$$

For instance, Henshall and Shallit ask

- *Can the techniques be applied to detect abelian powers in automatic sequences?*

Remark

The Thue–Morse word is abelian periodic, $\mathbf{t} \in \{ab, ba\}^\omega$, therefore abelian equivalence is "easy", but then problems occur for 2-abelian equivalence.

- J. Berstel, M. Crochemore, J.-E. Pin, Thue–Morse sequence and p-adic topology for the free monoid, *Disc. Math.* **76** (1989), 89–94.