Counting factors of automatic sequences up to abelian equivalence

Aline Parreau Michel Rigo Eric Rowland Élise Vandomme

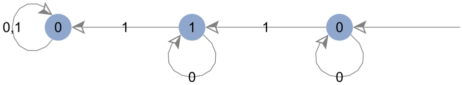
Université de Liège

2014 October 22

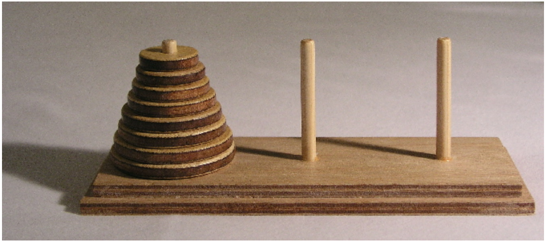
Automatic sequences

A sequence $(a_n)_{n \geq 0}$ is 2-automatic if there is DFAO whose output is a_n when fed the base-2 digits of n.

- Characteristic sequence $011010001 \cdots$ of powers of 2:

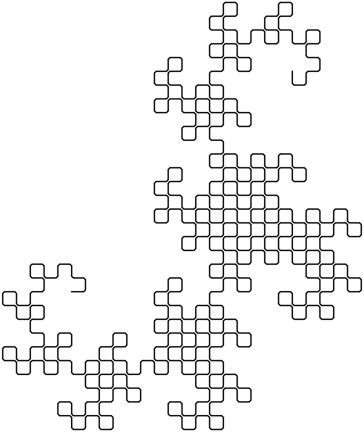


- Minimal solution to the "infinite" tower of Hanoi puzzle



Paper-folding sequence

$$LLRLLRRLLLRRLRRLLLRLLRRRLLRRLRRL\cdots$$

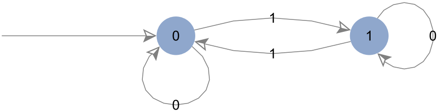


Let $T(n) =$ (number of 1s in the binary representation of n) mod 2.

The Thue–Morse sequence

$$\mathbf{t} = T(n)_{n \geq 0} = 01101001100101101001011001101001 \cdots$$

is 2-automatic. It is also cube-free.



$\mathbf{t} = \varphi^{\infty}(0)$ is a fixed point of the morphism $\varphi : 0 \to 01, 1 \to 10$.

A sequence is 2-automatic if and only if it is the image, under a coding, of a fixed point of a 2-uniform morphism (Cobham 1972).

Complexity of sequences

Given a sequence $\mathbf{x}$, what is its complexity?

Different measures of complexity:

- factor complexity $\mathcal{P}_{\mathbf{x}}(n)$: number of distinct factors of length n.
- abelian complexity $\mathcal{P}_{\mathbf{x}}^{ab}(n)$: number of factors of length n up to abelian equivalence (e.g., $001100 \equiv_{ab} 010010$).

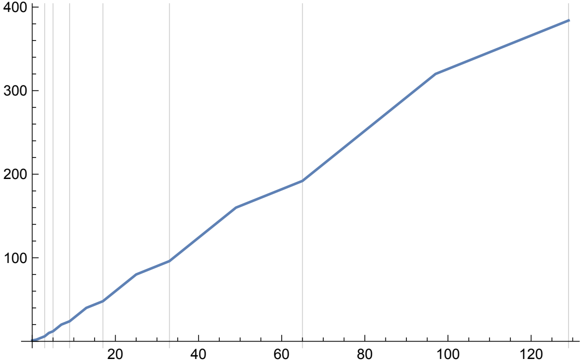
In general, $\mathcal{P}_{\mathbf{x}}^{ab}(n) \leq \mathcal{P}_{\mathbf{x}}(n)$.

Factor complexity of Thue–Morse

Theorem (Brlek 1989)

For $n \geq 3$, the factor complexity of the Thue–Morse sequence is

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



Abelian complexity of Thue–Morse

$$\begin{aligned}\mathbf{t} = \varphi^\infty(0) &= 0110100110010110\cdots \\ &= (01)(10)(10)(01)(10)(01)(01)(10)\cdots\end{aligned}$$

The abelian complexity of Thue–Morse is simpler, since $\varphi(0) \equiv_{ab} \varphi(1)$.

Proposition

For $n \geq 1$,

$$\mathcal{P}^{ab}_{\mathbf{t}}(n) = \begin{cases} 3 & \textit{if } n \textit{ is even} \\ 2 & \textit{if } n \textit{ is odd.} \end{cases}$$

ℓ-abelian equivalence

Definition

Two words u and v are ℓ-abelian equivalent, denoted $u \equiv_\ell v$, if $|u|_w = |v|_w$ for all words w of length $\leq \ell$.

Example

Let $u = 011010011$ and $v = 001101101$.

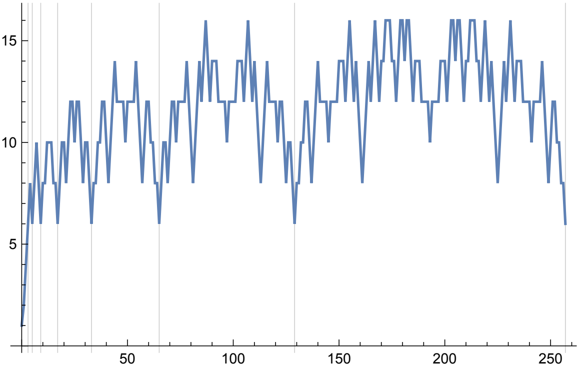
- $u \equiv_2 v$ because
  - $|u|_0 = 4 = |v|_0$, $|u|_1 = 5 = |v|_1$ and
  - $|u|_{00} = 1 = |v|_{00}$, $|u|_{01} = 3 = |v|_{01}$, etc.
- $u \not\equiv_3 v$ because $|u|_{101} = 1 \neq 2 = |v|_{101}$

But $01201011 \equiv_3 01012011$.

The ℓ-abelian complexity $\mathcal{P}_{\mathbf{x}}^{(\ell)}(n)$ of a sequence $\mathbf{x}$ is the number of factors of length n up to ℓ-abelian equivalence.

$$\mathcal{P}_{\mathbf{x}}^{ab}(n) = \mathcal{P}_{\mathbf{x}}^{(1)}(n) \leq \mathcal{P}_{\mathbf{x}}^{(\ell)}(n) \leq \mathcal{P}_{\mathbf{x}}^{(\infty)}(n) = \mathcal{P}_{\mathbf{x}}(n).$$

2-abelian complexity of Thue–Morse



Not piecewise linear like $\mathcal{P}_{\mathbf{t}}(n)$; not eventually periodic like $\mathcal{P}_{\mathbf{t}}^{ab}(n)$.

Karhumäki–Saarela–Zamboni 2013: $\mathcal{P}_{\mathbf{t}}^{(2)}(n) = O(\log n)$.

How to explain the nested structure?

Nested structure in automatic sequences

Definition

Let $k \geq 2$. The *k-kernel* of a sequence $\mathbf{x}$ is the set of sequences

$$\mathcal{K}_{k,\mathbf{x}} = \{(\mathbf{x}_{k^e n+r})_{n \geq 0} \mid e \geq 0,\ 0 \leq r \leq k^e - 1\}.$$

2-kernel of the Thue–Morse sequence:

$$\begin{array}{rrcll} & \mathbf{t} & = & 01101001100101101001011001101\cdots & \\ e=1, r=0 & (\mathbf{t}_{2n})_{n\geq 0} & = & 01101001100101101001011001101\cdots & = \mathbf{t} \\ e=1, r=1 & (\mathbf{t}_{2n+1})_{n\geq 0} & = & 10010110011010010110100110010\cdots & = \bar{\mathbf{t}} \end{array}$$

$$\mathcal{K}_{2,\mathbf{t}} = \{\mathbf{t}, \bar{\mathbf{t}}\}$$

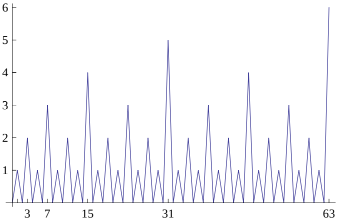
A sequence is 2-automatic if and only if its 2-kernel is finite.

Nested structure in integer sequences

Let $\nu_2(n)$ be the exponent of the largest power of 2 dividing n.

The "ruler sequence" $\nu_2(n+1)_{n\geq 0}$ is

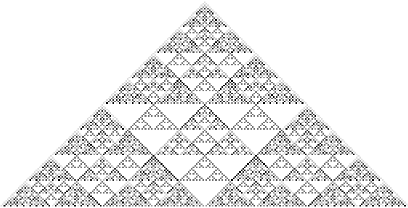
$$0\,1\,0\,2\,0\,1\,0\,3\,0\,1\,0\,2\,0\,1\,0\,4\,0\,1\,0\,2\,0\,1\,0\,3\,0\,1\,0\,2\,0\,1\,0\,5\,\cdots.$$

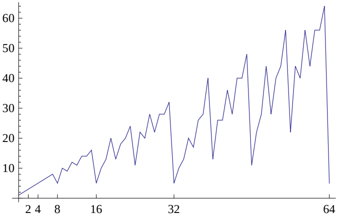


$$\nu_2(2n+0) = 1 + \nu_2(n)$$
$$\nu_2(2n+1) = 0$$

Counting nonzero binomial coefficients modulo 8

Let $s(n) = |\{0 \leq m \leq n : \binom{n}{m} \not\equiv 0 \mod 8\}|$.





1 2 3 4 5 6 7 8 5 10 9 12 11 14 14 16 5 10 13 20 13 18 20 24 $\cdots$

$$\begin{aligned} s(2n+1) &= 2s(n) \\ s(4n+0) &= s(2n) \\ s(8n+2) &= -2s(n) + 2s(2n) + s(4n+2) \\ s(8n+6) &= 2s(4n+2) \end{aligned}$$

Regular sequences

Definition (Allouche–Shallit 1992)

An integer sequence $s(n)_{n\geq 0}$ is *k-regular* if the $\mathbb{Z}$-module generated by the k-kernel

$$\{s(k^e n + r)_{n\geq 0} \mid e \geq 0,\ 0 \leq r \leq k^e - 1\}.$$

is finitely generated.

Theorem (Mossé 1996, Charlier–Rampersad–Shallit 2012)

The factor complexity of a k-automatic sequence is k-regular.

Question

Is the ℓ-abelian complexity of a k-automatic sequence k-regular?

Complexity and regularity

Theorem (Madill–Rampersad 2013)

The abelian complexity of the paperfolding sequence is 2-regular.

Let $\psi(0) = 01$, $\psi(1) = 00$. The period-doubling sequence is

$$\mathbf{p} = \psi^\infty(0) = 0100010101000100010 \cdots .$$

Theorem (Karhumäki–Saarela–Zamboni 2013)

The abelian complexity of the period-doubling sequence is 2-regular.

For Thue–Morse $\mathbf{t}$...
The abelian complexity is 2-regular (since it is eventually periodic).
Is the 2-abelian complexity 2-regular?

Proving 2-regularity

Guess and prove relations among sequences in the 2-kernel.

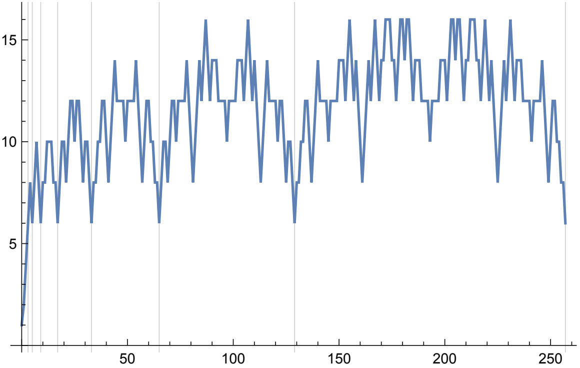
Let $\mathbf{x}_{2^e+r} = \mathcal{P}_{\mathbf{t}}^{(2)}(2^e n + r)$.

$$\begin{aligned}
\mathbf{x}_{5} &= \mathbf{x}_{3} \\
\mathbf{x}_{9} &= \mathbf{x}_{3} \\
\mathbf{x}_{12} &= -\mathbf{x}_{6} + \mathbf{x}_{7} + \mathbf{x}_{11} \\
\mathbf{x}_{13} &= \mathbf{x}_{7} \\
\mathbf{x}_{16} &= \mathbf{x}_{8} \\
\mathbf{x}_{17} &= \mathbf{x}_{3} \\
\mathbf{x}_{18} &= \mathbf{x}_{10} \\
\mathbf{x}_{20} &= -\mathbf{x}_{10} + \mathbf{x}_{11} + \mathbf{x}_{19} \\
\mathbf{x}_{21} &= \mathbf{x}_{11} \\
\mathbf{x}_{22} &= -\mathbf{x}_{3} - 2\mathbf{x}_{6} + \mathbf{x}_{7} + 3\mathbf{x}_{10} + \mathbf{x}_{11} - \mathbf{x}_{19} \\
\mathbf{x}_{23} &= -\mathbf{x}_{3} - 3\mathbf{x}_{6} + 2\mathbf{x}_{7} + 3\mathbf{x}_{10} + \mathbf{x}_{11} - \mathbf{x}_{19} \\
\mathbf{x}_{24} &= -\mathbf{x}_{3} + \mathbf{x}_{7} + \mathbf{x}_{10} \\
\mathbf{x}_{25} &= \mathbf{x}_{7} \\
\mathbf{x}_{26} &= -\mathbf{x}_{3} + \mathbf{x}_{7} + \mathbf{x}_{10} \\
\mathbf{x}_{27} &= -2\mathbf{x}_{3} + \mathbf{x}_{7} + 3\mathbf{x}_{10} - \mathbf{x}_{19} \\
\mathbf{x}_{28} &= -2\mathbf{x}_{3} + \mathbf{x}_{7} + 3\mathbf{x}_{10} - \mathbf{x}_{14} + \mathbf{x}_{15} - \mathbf{x}_{19} \\
\mathbf{x}_{29} &= \mathbf{x}_{15} \\
\mathbf{x}_{30} &= -\mathbf{x}_{3} + 3\mathbf{x}_{6} - \mathbf{x}_{7} - \mathbf{x}_{10} - \mathbf{x}_{11} + \mathbf{x}_{15} + \mathbf{x}_{19} \\
\mathbf{x}_{31} &= -3\mathbf{x}_{3} + 6\mathbf{x}_{6} - 2\mathbf{x}_{11} - 3\mathbf{x}_{14} + 2\mathbf{x}_{15} + \mathbf{x}_{19} \\
\mathbf{x}_{32} &= \mathbf{x}_{8} \\
\mathbf{x}_{33} &= \mathbf{x}_{3} \\
\mathbf{x}_{34} &= \mathbf{x}_{10} \\
\mathbf{x}_{35} &= \mathbf{x}_{11} \\
\mathbf{x}_{36} &= -\mathbf{x}_{10} + \mathbf{x}_{11} + \mathbf{x}_{19} \\
\mathbf{x}_{37} &= \mathbf{x}_{19} \\
\mathbf{x}_{38} &= -\mathbf{x}_{3} + \mathbf{x}_{10} + \mathbf{x}_{19}
\end{aligned}$$

$$\begin{aligned}
\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}$$

Guessed by Rigo–Vandomme, proved by Greinecker (2014).

More general approach?

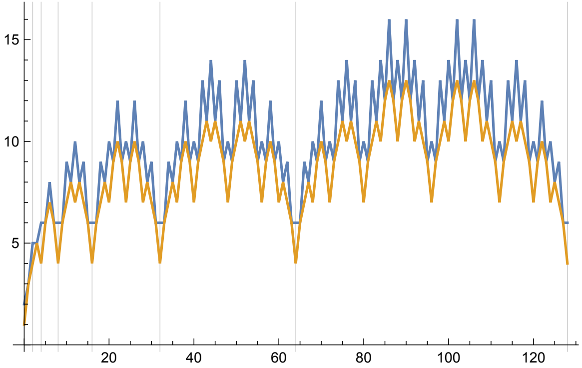


Apparent symmetry between powers of 2:
$\mathcal{P}_{\mathbf{t}}^{(2)}(2^{\ell+1} - r) = \mathcal{P}_{\mathbf{t}}^{(2)}(2^{\ell} + r)$.

And there is some relation between $\mathcal{P}_{\mathbf{t}}^{(2)}(2^{\ell} + r)$ and $\mathcal{P}_{\mathbf{t}}^{(2)}(r)$.

2-abelian complexity of the period-doubling sequence

Recall $\mathbf{p} = \psi^\infty(0) = 0100010101000100\cdots$ where $\psi(0) = 01$, $\psi(1) = 00$.

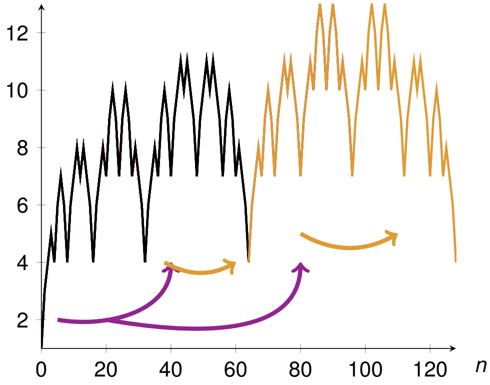


The 2-abelian complexity of $\mathbf{p}$ is closely related to the 1-abelian complexity of the 2-block coding of $\mathbf{p}$.

Reduction to 1-abelian complexity

The 2-block coding of **p** is the fixed point of $0 \to 12, 1 \to 12, 2 \to 00$:

$$\mathbf{x} = 1200121212001200120012121200121212 00 \cdots$$



Nice translation and reflection relations:

$$\mathcal{P}_{\mathbf{x}}^{(1)}(2^\ell + r) = \mathcal{P}_{\mathbf{x}}^{(1)}(r) + 3 \qquad \mathcal{P}_{\mathbf{x}}^{(1)}(2^{\ell+1} - r) = \mathcal{P}_{\mathbf{x}}^{(1)}(2^\ell + r)$$

From relations to regularity

Do these relations imply 2-regularity?
Not clear; our relations peel off the most significant digit, whereas 2-kernel relations should peel off the least significant digit.

Theorem

Suppose $s(n)_{n\geq 0}$ satisfies

$$s(2^\ell + r) = \begin{cases} s(r) + c & \text{if } r \leq 2^{\ell-1} \\ s(2^{\ell+1} - r) & \text{if } r > 2^{\ell-1} \end{cases}$$

for all $\ell \geq \ell_0$ and $0 \leq r \leq 2^\ell - 1$. Then $s(n)_{n\geq 0}$ is 2-regular.

Outline of the proof:

- Prove the case $c = 0$.
- Prove general 2-kernel relations by induction on $n = 2^\ell + r$.

Proving the translation and reflection relations

The 2-block coding of $\mathbf{p}$:

$$\mathbf{x} = 1200121212001200120012121200121212 00\cdots$$

1. Consider
$$\Delta_0(n) = \max_{|u|=n} |u|_0 - \min_{|u|=n} |u|_0.$$
2. $\Delta_0(n)$ is closely related to $\mathcal{P}^{(1)}_{\mathbf{x}}(n)$ since 1 and 2 alternate in $\mathbf{x}$.
3. Prove the translation and reflection relations for $\Delta_0(n)$.

Therefore $\mathcal{P}^{(1)}_{\mathbf{x}}(n)$ is 2-regular, and this implies the following.

Theorem

The 2-abelian complexity $\mathcal{P}^{(2)}_{\mathbf{p}}(n)$ of the period-doubling sequence is 2-regular.

Back to Thue–Morse

The 2-block coding of Thue–Morse

$$\mathbf{y} = 132120132012132120121320\cdots$$

is a fixed point of $0 \to 12$, $1 \to 13$, $2 \to 20$, $3 \to 21$.

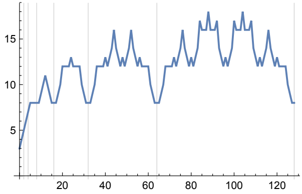
1. Consider the function $\Delta_{1,2}(n)$.
2. $\Delta_{1,2}(n)$ is closely related to $\mathcal{P}_{\mathbf{y}}^{(1)}(n)$ since $1, 2$ alternate and $0, 3$ alternate in $\mathbf{y}$.
3. Prove the translation and reflection relations for $\Delta_{1,2}(n)$.

Theorem

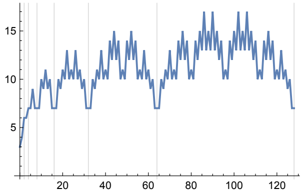
The 2-abelian complexity $\mathcal{P}_{\mathbf{t}}^{(2)}(n)$ of the Thue–Morse sequence is 2-regular.

Other sequences

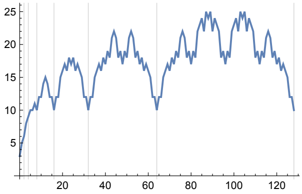
Some 2-abelian complexity sequences appear to satisfy the reflection:



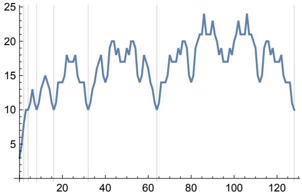
{0 → 01, 1 → 02, 2 → 01}



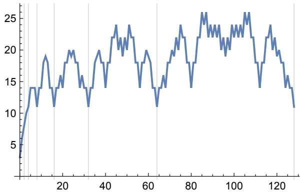
{0 → 01, 1 → 12, 2 → 11}



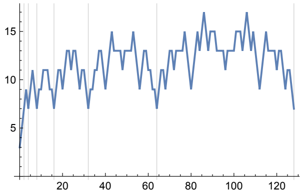
{0 → 01, 1 → 20, 2 → 01}



{0 → 01, 1 → 20, 2 → 10}



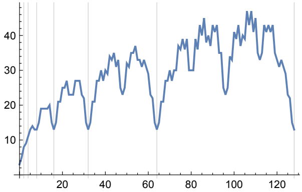
{0 → 01, 1 → 12, 2 → 01}



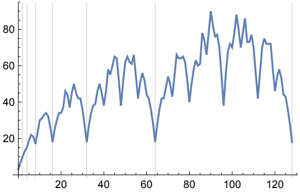
{0 → 01, 1 → 12, 2 → 21}

Other sequences

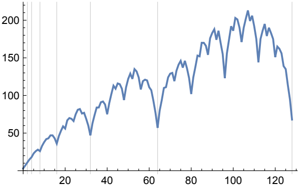
But not all . . .



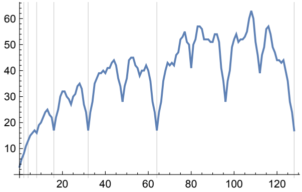
{0 → 01, 1 → 02, 2 → 00}



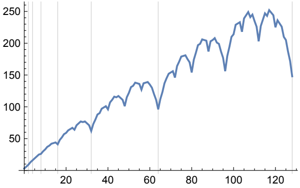
{0 → 01, 1 → 02, 2 → 12}



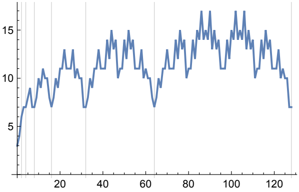
{0 → 01, 1 → 02, 2 → 21}



{0 → 01, 1 → 21, 2 → 00}



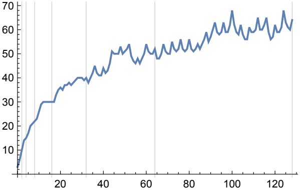
{0 → 01, 1 → 22, 2 → 00}



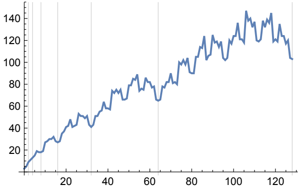
{0 → 01, 1 → 22, 2 → 12}

Other sequences

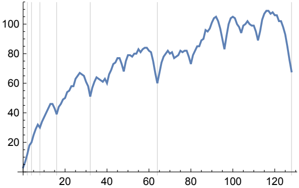
But not all . . .



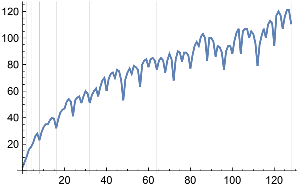
{0 → 01, 1 → 02, 2 → 10}



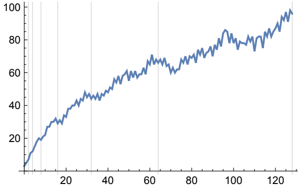
{0 → 01, 1 → 02, 2 → 11}



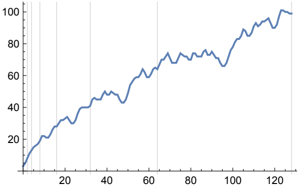
{0 → 01, 1 → 02, 2 → 20}



{0 → 01, 1 → 12, 2 → 02}



{0 → 01, 1 → 20, 2 → 00}



{0 → 01, 1 → 20, 2 → 22}