A new approach to the 2-regularity of the $\ell$-abelian complexity of 2-automatic sequences (extended abstract)

Aline Parreau*, Michel Rigo† Eric Rowland‡ and Élise Vandomme§

3 June 2014

Abstract

We show that a sequence satisfying a certain symmetry property is 2-regular in the sense of Allouche and Shallit. We apply this theorem to develop a general approach for studying the $\ell$-abelian complexity of 2-automatic sequences. In particular, we prove that the period-doubling word and the Thue–Morse word have 2-abelian complexity sequences that are 2-regular. Along the way, we also prove that the 2-block codings of these two words have 1-abelian complexity sequences that are 2-regular.

1 Introduction

This extended abstract is about some structural properties of integer sequences that occur naturally in combinatorics on words. Since the fundamental work of Cobham [6], the so-called automatic sequences have been extensively studied. We refer the reader to [3] for basic definitions and properties. These infinite words over a finite alphabet can be obtained by iterating a prolongable morphism of constant length to get an infinite word (and then, an extra letter-to-letter morphism, also called coding, may be applied once). As a fundamental example, the Thue–Morse word $t = \sigma(\omega)(0) = 0110100110010110 \cdots$ is a fixed point of the morphism $\sigma$ over the free monoid $\{0, 1\}^*$ defined by $\sigma(0) = 01$, $\sigma(1) = 10$. Similarly, the period-doubling word $p = \psi(\omega)(0) = 01000101010001000100 \cdots$ is a fixed point of the morphism $\psi$ over $\{0, 1\}^*$ defined by $\psi(0) = 01$, $\psi(1) = 00$.

Let $k \geq 2$ be an integer. One characterization of $k$-automatic sequences is that their $k$-kernels are finite; see [7] or [3, Section 6.6].

Definition 1. The $k$-kernel of a sequence $s = s(n)_{n \geq 0}$ is the set

$$K_k(s) = \{s(ik^n + j)_{n \geq 0} : i \geq 0 \text{ and } 0 \leq j < k^i\}.$$  

For instance, the 2-kernel $K_2(t)$ of the Thue–Morse word contains exactly two elements, namely $t$ and $\sigma^2(1)$.

A natural generalization of automatic sequences to sequences on an infinite alphabet is given by the notion of $k$-regular sequences. We will restrict ourselves to sequences taking integer values only.

---

*FNRS post-doctoral fellow at the University of Liege
†University of Liege
‡BelFPD-COFUND post-doctoral fellow at the University of Liege
§Corresponding author, University of Liege, University of Grenoble

1For the full version of this paper, see [15].
Definition 2. Let \( k \geq 2 \) be an integer. A sequence \( s = (s(n))_{n \geq 0} \in \mathbb{Z}^N \) is \( k \)-regular if \( (K_k(s)) \) is a finitely-generated \( \mathbb{Z} \)-module, i.e., there exist a finite number of sequences \( s_1(n)_{n \geq 0}, \ldots, s_{\ell}(n)_{n \geq 0} \) such that every sequence in the \( k \)-kernel \( K_k(s) \) is a \( \mathbb{Z} \)-linear combination of the \( s_r \)'s. Otherwise stated, for all \( i \geq 0 \) and for all \( j \in \{0, \ldots, k^i - 1\} \), there exist integers \( c_1, \ldots, c_{\ell} \) such that

\[
\forall n \geq 0, \quad s(k^i n + j) = \sum_{r=1}^{\ell} c_r t_r(n).
\]

Allouche and Shallit give many natural examples of \( k \)-regular sequences and classical results [11, 2]. The \( k \)-regularity of a sequence provides us with structural information about how the different terms are related to each other.

We will often make use of the following composition theorem for a function \( F \) defined piecewise on several \( k \)-automatic sets.

Lemma 3. Let \( k \geq 2 \). Let \( P_1, \ldots, P_\ell : \mathbb{N} \to \{0,1\} \) be unary predicates that are \( k \)-automatic. Let \( f_1, \ldots, f_\ell \) be \( k \)-regular functions. The function \( F : \mathbb{N} \to \mathbb{N} \) defined by

\[
F(n) = \sum_{i=1}^{\ell} f_i(n) P_i(n)
\]

is \( k \)-regular.

A classical measure of complexity of an infinite word \( x \) is its factor complexity \( P_x^{(\infty)} : \mathbb{N} \to \mathbb{N} \) which maps \( n \) to the number of distinct factors of length \( n \) occurring in \( x \). It is well known that a \( k \)-automatic sequence \( x \) has a \( k \)-regular factor complexity function [13, 6]. As an example, again for the Thue–Morse word, we have

\[
P_t^{(\infty)}(2n + 1) = 2P_t^{(\infty)}(n + 1) \quad \text{and} \quad P_t^{(\infty)}(2n) = P_t^{(\infty)}(n + 1) + P_t^{(\infty)}(n)
\]

for all \( n \geq 2 \).

Recently there has been a renewal of interest in abelian notions arising in combinatorics on words (e.g., avoiding abelian or \( \ell \)-abelian patterns, abelian bordered words, etc.). For instance, two finite words \( u \) and \( v \) are abelian equivalent if one is obtained by permuting the letters of the other one. Since the Thue–Morse word is an infinite concatenation of factors 01 and 10, this word is abelian periodic of period 2. The abelian complexity of an infinite word \( x \) is a function \( P_x^{(1)} : \mathbb{N} \to \mathbb{N} \) which maps \( n \) to the number of distinct factors of length \( n \) occurring in \( x \), counted up to abelian equivalence. Madill and Rampersad [12] provided the first example of regularity in this setting: the abelian complexity of the paper-folding word (which is another typical example of an automatic sequence) is unbounded and \( 2 \)-regular.

Let \( \ell \geq 1 \) be an integer. Based on [9] the notions of abelian equivalence and thus abelian complexity were recently extended to \( \ell \)-abelian equivalence and \( \ell \)-abelian complexity [10].

Definition 4. Let \( u, v \) be two finite words. We let \( |u|_v \) denote the number of occurrences of the factor \( v \) in \( u \). Two finite words \( x \) and \( y \) are \( \ell \)-abelian equivalent if \( |x|_v = |y|_v \) for all words \( v \) of length \( |v| \leq \ell \).

As an example, the words 011010011 and 001101011 are 2-abelian equivalent but not 3-abelian equivalent (the factor 010 occurs in the first word but not in the second one). Hence one can define the function \( P_x^{(\ell)} : \mathbb{N} \to \mathbb{N} \) which maps \( n \) to the number of distinct factors of length \( n \) occurring in the infinite word \( x \), counted up to \( \ell \)-abelian equivalence. In particular, for any infinite word \( x \), we have for all \( n \geq 0 \)

\[
P_x^{(1)}(n) \leq \cdots \leq P_x^{(\ell)}(n) \leq P_x^{(\ell+1)}(n) \leq \cdots \leq P_x^{(\infty)}(n).
\]
Since we are interested in $\ell$-abelian complexity, it is natural to consider the following operation that permits us to compare factors of length $\ell$ occurring in an infinite word.

**Definition 5.** Let $\ell \geq 1$. The $\ell$-block coding of the word $w = w_0w_1w_2 \cdots$ over the alphabet $A$ is the word
\[
\text{block}(w, \ell) = (w_0 \cdots w_{\ell-1}) (w_1 \cdots w_{\ell}) (w_2 \cdots w_{\ell+1}) \cdots (w_j \cdots w_{j+\ell-1}) \cdots
\]
over the alphabet $A^\ell$. If $A = \{0, \ldots, r - 1\}$, then it is convenient to identify $A^\ell$ with the set $\{0, \ldots, r^\ell - 1\}$ and each word $w_0 \cdots w_{\ell-1}$ of length $\ell$ is thus replaced with the integer obtained by reading the word in base $r$, i.e., $\sum_{i=0}^{\ell-1} w_i r^\ell - 1$. It is well known that the $\ell$-block coding of a $k$-automatic sequence is again a $k$-automatic sequence [2]. One can also define accordingly the $\ell$-block coding of a finite word $u$ of length at least $\ell$. For example, the $2$-block codings of $011010011$ and $001101101$ are respectively $13212013$ and $01321321$, which are abelian equivalent.

**Lemma 6.** [10, Lemma 2.3] Let $\ell \geq 1$. Two finite words $u$ and $v$ of length at least $\ell - 1$ are $\ell$-abelian equivalent if and only if they share the same prefix (resp. suffix) of length $\ell - 1$ and the words $\text{block}(u, \ell)$ and $\text{block}(v, \ell)$ are abelian equivalent.

In this paper, we show that both the period-doubling word $p$ and the Thue–Morse word $t$ have $2$-abelian complexity sequences which are $2$-regular. In [11], the authors studied the asymptotic behavior of $P_t^{(\ell)}(n)$ and also derived some recurrence relations showing that the abelian complexity $P_p^{(\ell)}(n)_{n \geq 0}$ of the period-doubling word $p$ is $2$-regular. From [1], one can deduce some other relations about the abelian complexity of $p$.

Given the first few terms of a sequence, the second and last authors conjectured the $2$-regularity of the sequence $P_t^{(2)}(n)_{n \geq 0}$ by exhibiting relations that should be satisfied (and proved some recurrence relations for this sequence) [16]. See [2, Section 6] for such a “predictive” algorithm that recognizes regularity. Recently, Greinecker proved the recurrence relations needed to prove the $2$-regularity of this sequence [8]. Hopefully, the two approaches are complementary: in this paper, we prove $2$-regularity without exhibiting the explicit recurrence relations.

Our approach is based on Theorem 7 which establishes the $2$-regularity of a large family of sequences satisfying a recurrence relation with a parameter $c$ and $2^{\ell_0}$ initial conditions. Computer experiments suggest that many $2$-abelian complexity functions satisfy such a reflection property.

**Theorem 7.** Let $\ell_0 \geq 0$ and $c \in \mathbb{Z}$. Suppose $s(n)_{n \geq 0}$ is a sequence such that, for all $\ell \geq \ell_0$ and $0 \leq r \leq 2^\ell - 1$, we have
\[
s(2^\ell + r) = \begin{cases} 
s(r) + c & \text{if } r \leq 2^\ell - 1 \\
s(2^\ell + 1 - r) & \text{if } r > 2^\ell - 1. \end{cases}
\]
(1)
Then $s(n)_{n \geq 0}$ is $2$-regular.

It turns out that the general solution of Equation (1) can be expressed naturally in terms of the sequence $A(n)_{n \geq 0}$ satisfying the recurrence for $\ell_0 = 0$ and $c = 1$ with $A(0) = 0$. The sequence $A(n)_{n \geq 0}$ appears as [11] A007302. Allouche and Shallit [2] identified this sequence as an example of a regular sequence.

From Equation (1) one can get some information about the asymptotic behavior of the sequence $s(n)_{n \geq 0}$. We have $s(n) = O(\log n)$, and moreover
\[
s \left( 2^{\ell+1}-1 \right) = s \left( 4^\ell + \cdots + 4^1 + 4^0 \right) = \left( \ell - \left\lfloor \frac{\ell_0 - 1}{2} \right\rfloor \right) c + \sum_{r=1}^{4^{\frac{(\ell_0 + 1)/2 - 1}}}{r}
\]
for $\ell \geq \left\lfloor \frac{\ell_0 - 1}{2} \right\rfloor$. At the same time, there are many subsequences of $s(n)_{n \geq 0}$ which are constant; for example, $s(2^\ell) = c$ for $\ell \geq \ell_0$.

**Example 8.** As an illustration of the reflection property described in Theorem 7, we consider in Figure 1 the abelian complexity of the $2$-block coding of the period-doubling word $p$. \

---


---


2-Abelian complexity of the period-doubling word

To show the 2-regularity of the 2-abelian complexity of $p$, we consider first the abelian complexity of the 2-block coding $x$ of $p$ and then we compare $P_x^{(1)}(n)$ with $P_p^{(2)}(n+1)$. The 2-block coding of $p$ is given by

$$x := \text{block}(p, 2) = \phi^{\omega}(1) = 12001212120012001200121212001212 \cdots$$

where $\phi$ is the morphism defined by $0 \mapsto 12$, $1 \mapsto 12$, $2 \mapsto 00$.

We introduce functions related to the number of 0’s in the factors of $x$ of length $n$. Let $n \in \mathbb{N}$. We let $\max_0(n)$ (resp. $\min_0(n)$) denote the maximum (resp. minimum) number of 0’s in a factor of $x$ of length $n$. Let $\Delta_0(n) := \max_0(n) - \min_0(n)$ be the difference between these two values.

To prove the 2-regularity of the sequence $P_x^{(1)}(n)_{n \geq 0}$, we first express $P_x^{(1)}(n)$ in terms of $\Delta_0(n)$.

**Proposition 9.** For $n \in \mathbb{N}$,

$$P_x^{(1)}(n) = \begin{cases} \frac{3}{2} \Delta_0(n) + \frac{3}{2} & \text{if } \Delta_0(n) \text{ is odd} \\ \frac{3}{2} \Delta_0(n) + 1 & \text{if } \Delta_0(n) \text{ and } n - \min_0(n) \text{ are even} \\ \frac{3}{2} \Delta_0(n) + 2 & \text{if } \Delta_0(n) \text{ and } n - \min_0(n) + 1 \text{ are even} \end{cases}$$

To be able to apply the composition result given by Lemma 3 to the expression of $P_x^{(1)}$, we have to prove that the sequence $\Delta_0(n)_{n \geq 0}$ is 2-regular (this is a consequence of the following result) and that the predicates occurring in the previous statement are 2-automatic.

**Proposition 10.** Let $\ell \geq 2$ and $0 \leq r < 2^\ell$. We have

$$\Delta_0(2^\ell + r) = \begin{cases} \Delta_0(r) + 2 & \text{if } r \leq 2^{\ell-1} \\ \Delta_0(2^{\ell+1} - r) & \text{if } r > 2^{\ell-1} \end{cases}$$

As a consequence of Propositions 9 and 10, $P_x^{(1)}(n)$ satisfies a reflection recurrence as in Theorem 7 with $\ell_0 = 2$ and $c = 3$. This implies again that the sequence is 2-regular.

Now consider the 2-abelian complexity $P_p^{(2)}$. To apply Lemma 3, we will express $P_p^{(2)}$ in terms of the abelian complexity $P_x^{(1)}$ and the following additional 2-automatic functions.

**Definition 11.** We define the max-jump function $\text{MJ}_0 : \mathbb{N} \to \{0, 1\}$ by $\text{MJ}_0(n) = 1$ when the function $\max_0(n)$ increases. Similarly, let $\text{mj}_0 : \mathbb{N} \to \{0, 1\}$ be the min-jump function defined by $\text{mj}_0(n) = \min_0(n+1) - \min_0(n)$.

To compute $P_p^{(2)}$, we will study when an abelian equivalence class of $x$ splits into two 2-abelian equivalence classes of $p$. Let $\mathcal{X}$ be an abelian equivalence class of factors of $x$ of length $n$ with
Proposition 13. Let \( n \geq 1 \) be an integer. Then
\[
\mathcal{P}_p^{(2)}(n + 1) - \mathcal{P}_x^{(1)}(n) = \begin{cases} 
0 & \text{if } n \text{ is odd} \\
\Delta(n) - 2 + 1 - \min_{\text{odd}}(n) & \text{if } n \text{ is even.}
\end{cases}
\]

In particular, the sequence \( \mathcal{P}_p^{(2)}(n)_{n \geq 0} \) is 2-regular.

### 3 2-Abelian complexity of the Thue–Morse word

In this section, we turn our attention to the Thue–Morse word \( t \). The approach here is similar to the one of the period-doubling word: we consider the abelian complexity of \( y = \text{block}(t, 2) \), and then we compare \( \mathcal{P}_y^{(1)}(n) \) with \( \mathcal{P}_x^{(2)}(n + 1) \). The 2-block coding of \( t \) is given by
\[
y := \text{block}(t, 2) = \nu^\omega(1) = 132120132012132120121320 \cdots
\]
where \( \nu \) is the morphism defined by \( \nu : 0 \mapsto 12, 1 \mapsto 13, 2 \mapsto 20, 3 \mapsto 21 \).

For the Thue–Morse word, the appropriate statistic for factors of \( y \) is the total number of 1's and 2's (or, equivalently, the total number of 0's and 3's). Therefore, for \( n \in \mathbb{N} \) we set
\[
\Delta_1(n) := \max_{\text{odd}}(n) - \min_{\text{odd}}(n) \quad \text{where } \max_{\text{odd}}(n) \text{ (resp. } \min_{\text{odd}}(n) \text{)} \text{ denote the maximum (resp. minimum) of } \{|u|_1 + |u|_2 : u \text{ is a factor of } y \text{ with } |u| = n\}.
\]

In particular, \( \Delta_1(n) + 1 \) is the abelian complexity function \( \mathcal{P}_p^{(1)}(n) \) of the period-doubling word. This function was also studied in [4, 11]. Here we can obtain relations for \( \Delta_1 \) of the same type as in Theorem 7.

As in the previous section, the fact that \( \mathcal{P}_y^{(1)}(n)_{n \geq 0} \) is 2-regular will follow from Lemma 4 applied to the next statement.

Proposition 14. Let \( n \in \mathbb{N} \). We have
\[
\mathcal{P}_y^{(1)}(n) = \begin{cases} 
2\Delta_1(n) + 2 & \text{if } n \text{ is odd} \\
\frac{3}{2}\Delta_1(n) + \frac{5}{2} & \text{if } n \text{ and } \Delta_1(n) + 1 \text{ are even} \\
\frac{5}{2}\Delta_1(n) + 4 + \min_1(\Delta_1(n) + 1) & \text{if } n, \Delta_1(n) \text{ and } \min_1(\Delta_1(n) + 1) \text{ are even}
\end{cases}
\]

As in Section 2 we define two new functions \( \max_{\text{odd}}(n) \) and \( \min_{\text{odd}}(n) \) analogously to Definition 11.

This permits us to compute the difference \( \mathcal{P}_x^{(2)}(n + 1) - \mathcal{P}_y^{(1)}(n) \).

Theorem 15. Let \( n \in \mathbb{N} \). The difference \( \mathcal{P}_x^{(2)}(n + 1) - \mathcal{P}_y^{(1)}(n) \) is equal to
\[
\begin{align*}
\Delta_1(n) + 2 & - 2 \max_{\text{odd}}(n) - 2 \min_{\text{odd}}(n) & \text{if } n, \min_1(\Delta_1(n) + 1) \text{ and } \max_1(\Delta_1(n) + 1) \text{ are odd} \\
\Delta_1(n) + 1 & - 2 \max_{\text{odd}}(n) & \text{if } n, \min_1(\Delta_1(n) + 1) \text{ and } \max_1(\Delta_1(n) + 1) \text{ are odd} \\
\Delta_1(n) + 1 & - 2 \min_{\text{odd}}(n) & \text{if } n, \max_1(\Delta_1(n) + 1) \text{ and } \min_1(\Delta_1(n) + 1) \text{ are odd} \\
\Delta_1(n) + 1 & - \frac{1}{2} \Delta_1(n) + 1 & \text{if } n, \min_1(\Delta_1(n) + 1) \text{ and } \max_1(\Delta_1(n) + 1) \text{ are even} \\
\frac{1}{2} \Delta_1(n) + 1 & & \text{if } n, \max_1(\Delta_1(n) + 1) \text{ and } \min_1(\Delta_1(n) + 1) \text{ are even}
\end{align*}
\]

In particular, the sequence \( \mathcal{P}_x^{(2)}(n)_{n \geq 0} \) is 2-regular.
4 Conclusions

The two examples treated in this paper suggest that a general framework to study the $\ell$-abelian complexity of $k$-automatic sequences may exist. Indeed, one conjectures that any $k$-automatic sequence has an $\ell$-abelian complexity function that is $k$-regular. As an example, if we consider the 3-block coding of the period-doubling word,

$$z = \text{block}(p, 3) = 2401252401240125240124 \cdots.$$ 

The abelian complexity $P_z^{(1)}(n)_{n \geq 0} = (1, 5, 5, 8, 6, 10, 19, 11, \ldots)$ seems to satisfy, for $\ell \geq 4$, the following relations (which are quite similar to what we have discussed so far)

$$P_z^{(1)}(2^\ell + r) = \begin{cases} 
P_z^{(1)}(r) + 5 & \text{if } r \leq 2^{\ell-1} \text{ and } r \text{ even} \\
P_z^{(1)}(r) + 7 & \text{if } r \leq 2^{\ell-1} \text{ and } r \text{ odd} \\
P_z^{(1)}(2^{\ell+1} - r) & \text{if } r > 2^{\ell-1}.
\end{cases}$$

Then, the second step would be to relate $P_p^{(3)}$ with $P_z^{(1)}$.

References


[4] Francine Blanchet-Sadri and James D. Currie and Narad Rampersad and Nathan Fox, Abelian complexity of fixed point of morphism $0 \mapsto 012, 1 \mapsto 02, 2 \mapsto 1$, *INTEGERS* **14** (2014) #A11.


[12] Blake Madill and Narad Rampersad, The abelian complexity of the paperfolding word

[13] Brigitte Mossé, Reconnaissabilité des substitutions et complexité des suites automatiques,


[15] Aline Parreau, Michel Rigo, Eric Rowland and Élise Vandomme, A new approach to the

[16] Michel Rigo and Élise Vandomme, 2-abelian complexity of the Thue–Morse sequence,
http://hdl.handle.net/2268/135841, December 2012.