Information and Computation 288 (2022) 104710

Contents lists available at ScienceDirect

Information and Computation

www.elsevier.com/locate/yinco

Complexity of automatic sequences

Hans Zantema [b,a,*], Wieb Bosma [a]

[a] *Radboud University Nijmegen, the Netherlands*
[b] *Eindhoven University of Technology, the Netherlands*

ARTICLE INFO

Article history:
Received 16 June 2020
Received in revised form 18 December 2020
Accepted 13 January 2021
Available online 10 February 2021

ABSTRACT

Automatic sequences can be defined by DFAs with output (DFAO) in two natural ways. We propose to consider the minimal size of a corresponding DFAO as the complexity measure of the automatic sequence, for both variants. This paper compares these complexity measures and investigates their properties, such as the relationships with kernel and morphic sequences. There exist automatic sequences for which the one complexity is exponentially greater than the other one, in both directions. For both complexity measures we investigate the effect of taking basic operations on sequences, like removing or adding an initial element, combining sequences, or taking arithmetic subsequences, and observe that these operations may increase the complexity at most polynomially. For periodic sequences we give sharp bounds for both complexity measures.

1. Introduction

Automatic sequences form an important class of infinite sequences over a finite alphabet; roughly speaking it is a first regular class going beyond ultimately periodic sequences. They have been extensively studied, in particular in the book [1] by Allouche and Shallit that serves as the main reference for research in this area. More recent references on the topic include [12,7].

Automatic sequences depend on a base $k > 1$, with special interest for $k = 2$. Two well-known 2-automatic sequences are the Thue-Morse sequence and the regular paper folding sequence, to be defined in Section 2. Automatic sequences admit several equivalent characterizations, many of which are closely related to the following two. In the first one the ith element a_i of the sequence a is the output of a DFAO when taking as input the k-ary expansion of i. The second one is similar, but then the reverse of the k-ary expansion of i is taken as input. It is natural to consider the minimal size of a corresponding DFAO as the complexity measure of the automatic sequence, for both variants, and we denote them by $\|a\|_k$ and $\|a\|_k^R$. These complexity measures are the main topic of this paper, they differ from other complexity measures like the one from [11]. We show how our complexity measures relate to other characterizations of automatic sequences; in particular, $\|a\|_k^R$ is closely related to the size of the *kernel* of a, and $\|a\|_k$ is closely related to the size of the smallest alphabet needed to describe a as a *morphic* sequence with respect to a k-uniform morphism. In doing so, we follow constructions as presented in [1], for which we investigate the precise effect on the measures $\|a\|_k$ and $\|a\|_k^R$.

A first result states that there is an exponential gap between both measures: there exist families of automatic sequences a for which $\|a\|_k^R$ is exponential in $\|a\|_k$, and families for which $\|a\|_k$ is exponential in $\|a\|_k^R$.

* Corresponding author.
E-mail addresses: h.zantema@tue.nl (H. Zantema), bosma@math.ru.nl (W. Bosma).

https://doi.org/10.1016/j.ic.2021.104710

A next natural question is about the effect of taking basic operations on sequences. For instance, for any sequence a its tail $\mathsf{tail}(a)$ is obtained by removing its first element. We show that $\|\mathsf{tail}(a)\|_k^R \leq 2\|a\|_k^R$ and $\|\mathsf{tail}(a)\|_k \leq (\|a\|_k)^2$ for all k-automatic sequences, and that the last inequality is sharp. Similar results hold for adding an initial element rather than removing it. Other operations are also considered, such as pointwise combining two sequences and taking arithmetic subsequences. The main observation about all of these basic operations f is that their sizes do not increase more than quadratically: $\|f(a)\|_k \leq (\|a\|_k)^2$ and $\|f(a)\|_k^R \leq (\|a\|_k^R)^2$ for all a. For cases where the bounds are quadratic, we show that these are sharp.

Another interesting question is what happens for periodic sequences. We derive a quadratic upper bound for $\|\cdot\|_k^R$ and a linear upper bound for $\|\cdot\|_k$, so opposite to the effect of tail. We also show that these upper bounds are reached if k and the period are relatively prime.

This paper can be seen as an extension of the paper [14] at LATA2020,[1] where it was honored by the best paper award. Compared to [14], the paper more than doubled in size. A main extension is the extensive combinatorial analysis of periodic sequences by the second author (Section 8). Other new contributions include a precise analysis of the length of the initial part of an automatic sequence to be inspected to determine the precise values of the complexity measures (Theorem 8), and a bound for arithmetic subsequences (Theorem 20).

This paper is organized as follows. In Section 2 we give the basic definitions and present two standard examples of 2-automatic sequences: the Thue-Morse sequence and the paper folding sequence, that will serve as the basis of many examples throughout the paper. In Section 3 we present our general techniques to compute the complexity in specific cases: upper bounds typically are given by constructions of automata, and we give a general lemma for proving lower bounds. Moreover, we discuss how SAT/SMT solving serves for computing the complexity measures of specific sequences. In Section 4 we investigate the exponential gap between $\|\cdot\|_k$ and $\|\cdot\|_k^R$. In Section 5 we define the kernel of an automatic sequence and investigate its relationship with $\|\cdot\|_k^R$. In Section 6 we present how to define automatic sequences as morphic sequences with respect to uniform morphisms, and investigate the relationship with $\|\cdot\|_k$. In Section 7 we investigate the effect of the basic operations mentioned above on the complexity measures. In Section 8 we give the bounds for both complexity measures for periodic sequences and show that they are sharp. We conclude in Section 9.

2. Basic definitions

Let $k \geq 2$ and $\Sigma_k = \{0, 1, \ldots, k-1\}$.

The set of infinite sequences $a = a_0a_1a_2a_3\cdots$ over a finite alphabet Σ is denoted by $\Sigma^{\mathbb{N}}$.

A DFA M with output (DFAO) is a tuple $M = (Q, \Sigma, \delta, q_0, \Gamma, \tau)$, where

- Q is the finite set of states,
- Σ is the finite input alphabet,
- $\delta : Q \times \Sigma \to Q$ is the transition function,
- $q_0 \in Q$ is the initial state,
- Γ is the finite output alphabet,
- $\tau : Q \to \Gamma$ is the output function.

DFAOs are depicted as arrows between states just as is usual for DFAs; the extra information that $\tau(q) = x$ is denoted by writing q/x in the state q.

As in DFAs, δ extends to $\delta : Q \times \Sigma^* \to Q$ by $\delta(q, \epsilon) = q$, $\delta(q, xu) = \delta(\delta(q, x), u)$. A DFAO M defines a function $f_M : \Sigma^* \to \Gamma$ defined by $f_M(u) = \tau(\delta(q_0, u))$. A function $f : \Sigma^* \to \Gamma$ is called a *finite state function* if a DFAO M exists such that $f = f_M$. For every finite state function f there exists a unique (up to renaming of states) DFAO M with a minimal number of states such that $f = f_M$.

A DFAO of which the input alphabet Σ is equal to $\Sigma_k = \{0, 1, \ldots, k-1\}$, is called a k-DFAO.

Every natural number n has a unique representation $(n)_k \in \Sigma_k^*$, where $(0)_k = \epsilon$ and

$$(n)_k = d_0d_1\cdots d_r \iff n = d_0k^r + d_1k^{r-1} + \cdots + d_{r-1}k + d_r \wedge d_0 > 0$$

for $n > 0$. So $(0)_2 = \epsilon$ and $(13)_2 = 1101$, for the decimal number $13 = (13)_{10}$. Note that non-empty strings of which the leftmost symbol is 0 do not occur as $(n)_k$ for any number n.

Conversely, every $u \in \Sigma_k^*$ represents a number $[u]_k$:

$$[d_0d_1\cdots d_r]_k = d_0k^r + d_1k^{r-1} + \cdots + d_{r-1}k + d_r.$$

For any Σ and any string $u \in \Sigma^*$ the reverse is $(u_1u_2\cdots u_n)^R = u_nu_{n-1}\cdots u_1$.

An infinite sequence $a \in \Gamma^{\mathbb{N}}$ is called k-automatic if a k-DFAO $M = (Q, \Sigma_k, \delta, q_0, \Gamma, \tau)$ exists such that $a_{[w]_k} = \tau(\delta(q_0, w))$ for all $w \in \Sigma_k^*$. According to Theorem 5.2.1 from [1] a is k-automatic if and only if a k-DFAO $M =$

[1] Although the actual conference was postponed due to Covid-19, the proceedings appeared properly.

$(Q_M, \Sigma_k, \delta_M, q_0, \Gamma, \tau_M)$ exists such that $\tau_M(\delta_M(q_0, (i)_k)) = a_i$ for all $i \in \mathbb{N}$. According to Theorem 5.2.3 from [1] a is k-automatic if and only if a k-DFAO $M = (Q_M, \Sigma_k, \delta_M, q_0, \Gamma, \tau_M)$ exists such that $\tau_M(\delta_M(q_0, (i)_k^R)) = a_i$ for all $i \in \mathbb{N}$.

Now we are ready to define the two natural measures $\|.\|_k$, $\|.\|_k^R$ for k-automatic sequences that we investigate in this paper.

Definition 1. For any k-automatic sequence $a = a_0a_1a_2a_3\cdots$ its

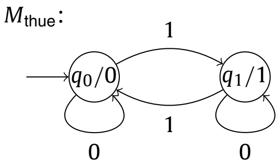
- size $\|a\|_k$ is the size of a smallest k-DFAO $M = (Q_M, \Sigma_k, \delta_M, q_0, \Gamma, \tau_M)$ such that $\tau_M(\delta_M(q_0, (i)_k)) = a_i$ for all $i \in \mathbb{N}$,
- reversed size $\|a\|_k^R$ is the size of a smallest k-DFAO $M = (Q_M, \Sigma_k, \delta_M, q_0, \Gamma, \tau_M)$ such that $\tau_M(\delta_M(q_0, (i)_k^R)) = a_i$ for all $i \in \mathbb{N}$.

Conversely, every k-DFAO $M = (Q_M, \Sigma_k, \delta_M, q_0, \Gamma, \tau_M)$ defines two infinite sequences $\mathsf{seq}_k(M)$ and $\mathsf{seq}_k^R(M)$ over Γ:

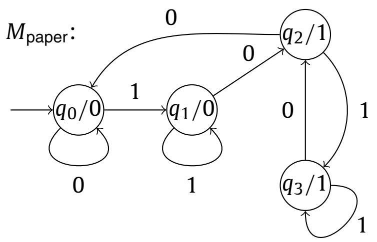
$$\mathsf{seq}_k(M)_i = \tau_M(\delta_M(q_0, (i)_k)) \text{ and } \mathsf{seq}_k^R(M)_i = \tau_M(\delta_M(q_0, (i)_k^R))$$

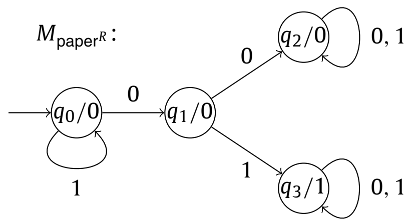
for all $i \in \mathbb{N}$. From the above definition it is immediate that $\|\mathsf{seq}_k(M)\|_k \leq |Q_M|$ and $\|\mathsf{seq}_k^R(M)\|_k^R \leq |Q_M|$.

Example 2. The *Thue-Morse sequence* $\mathsf{thue} = 0110100110010110\cdots$ is defined by $\mathsf{thue}_i = 0$ if the number of 1s in $(i)_2$ is even, and $\mathsf{thue}_i = 1$ if the number of 1s in $(i)_2$ is odd, see, e.g., [1] Section 1.6, or OEIS A010060. We have $\|\mathsf{thue}\|_2 = \|\mathsf{thue}\|_2^R = 2$, both realized by the DFAO below.



Example 3. The *regular paper-folding sequence* $\mathsf{paper} = 001001100011011\cdots$, also called *dragon curve sequence*, is defined by $\mathsf{paper}_i = m \bmod 2$ for every $i \geq 0$ for the unique representation $i = (2m+1)2^j - 1$, see, e.g., [1] Example 5.16., or OEIS A014577. We have $\|\mathsf{paper}\|_2 = \|\mathsf{paper}\|_2^R = 4$, respectively realized by the following two DFAOs.





The sequences thue and paper serve as leading examples throughout the paper.

3. Computing complexity

In this section we investigate techniques for bounding and computing $\|a\|_k$ and $\|a\|_k^R$ for specific sequences a. Throughout the paper we give several theorems providing upper bounds for specific operations. All of them are constructive: the bound is obtained by defining a DFAO and proving that it describes the intended sequence. If the DFAO is minimal, that is, all states are reachable and no two states are equivalent, then the size of the DFAO is not only an upper bound, but gives the exact value. Here two states q, q' are equivalent if $\tau(\delta(q,u)) = \tau(\delta(q',u))$ for all $u \in \Sigma^*$. This observation is similar to the well-known Myhill-Nerode theorem for DFAs. For our application a subtle complication is that we do not need $\tau(\delta(q,u)) = \tau(\delta(q',u))$ for all $u \in \Sigma^*$, but only for u that correspond to $(i)_k$ or $(i)_k^R$ for some $i \in \mathbb{N}$, so do not start or end in 0. This has been elaborated in [13].

The following lemma is a basic tool for proving lower bounds on $\|a\|_k$ and $\|a\|_k^R$.

Lemma 4. *Let a be a k-automatic sequence, and $m_1, \ldots, m_n \in \mathbb{N}$ such that for every $i \neq j$ there exists $v_{i,j} \in \Sigma_k^*$ satisfying $a_{[(m_i)_k v_{i,j}]_k} \neq a_{[(m_j)_k v_{i,j}]_k}$, then $\|a\|_k \geq n$.*

Let a be a k-automatic sequence, and $m_1, \ldots, m_n \in \mathbb{N}$ such that for every $i \neq j$ there exists $v_{i,j} \in \Sigma_k^*$ satisfying $a_{[v_{i,j}(m_i)_k]_k} \neq a_{[v_{i,j}(m_j)_k]_k}$, then $\|a\|_k^R \geq n$.

Proof. For the first claim let $M = (Q_M, \Sigma_k, \delta_M, q_0, \Gamma, \tau_M)$ be a smallest k-DFAO such that $\tau_M(\delta_M(q_0, (i)_k)) = a_i$ for all $i \in \mathbb{N}$. For $i = 1, 2, \ldots, n$ define $q_i = \delta_M(q_0, (m_i)_k)$. For $i \neq j$ from the assumption we obtain $\tau_M(\delta_M(q_i, v_{i,j})) \neq \tau_M(\delta_M(q_j, v_{i,j}))$, so $q_i \neq q_j$. This shows $|Q| \geq n$, so $\|a\|_k \geq n$.

The proof of the second claim is similar. □

A direct application of this lemma is a lower bound for periodic sequences.

Definition 5. A sequence a is *periodic* if it is of the shape $a = u^\omega = uuuuu\cdots$. It has *period* n if $n = |u|$ and u cannot be written as $u = v^i$ for some $i > 1$.

It is well-known ([1], Theorem 5.4.2) that every periodic sequence is k-automatic.

Corollary 6. *Let a be a periodic sequence of period n, for which k, n are relatively prime. Then $\|a\|_k \geq n$ and $\|a\|_k^R \geq n$.*

Proof. Let $a = u^\omega$ for $u = u_0 u_1 \cdots u_{n-1}$. Note that $a_i = u_{i \bmod n}$ for every $i \geq 0$. If for $0 \leq i < j < n$ we would have $u_{(i+x) \bmod n} = u_{(j+x) \bmod n}$ this would contradict the assumption that a has period n, so we conclude that

$$\forall\, 0 \leq i < j < n : \exists\, 0 \leq x < n : u_{(i+x) \bmod n} \neq u_{(j+x) \bmod n} \qquad (*)$$

For proving $\|a\|_k \geq n$ by Lemma 4 we choose $m_1, \ldots, m_n$ to be numbers that are all distinct modulo n. Choose p such that $n < k^p$. For x obtained from property $(*)$ we choose $v = k^p + x$, by which $a_{[(m_i)_k v]_k} = a_{m_i k^{p+1} + x}$. Since $m_1, \ldots, m_n$ are all distinct modulo n and n, k are relatively prime, also the n numbers $m_1 k^{p+1}, m_2 k^{p+1}, \ldots, m_n k^{p+1}$ are distinct modulo n. Now the condition of Lemma 4 holds by the property $(*)$, proving $\|a\|_k \geq n$.

The proof of $\|a\|_k^R \geq n$ by Lemma 4 is similar: then we choose $m_1, \ldots, m_n$ all having the same length in k-ary notation. □

In Section 8 we will investigate periodic sequences in more detail.

To apply Lemma 4 for proving a lower bound for $\|a\|_k$ or $\|a\|_k^R$ one has to choose specific numbers $m_1, \ldots, m_n$ and prove corresponding properties, which does not serve for automation. A completely different approach to prove lower bounds for $\|a\|_k$ and $\|a\|_k^R$ is the following. By definition $\|a\|_k > n$ means that no k-DFAO $(Q, \Sigma_k, \delta, q_0, \Gamma, \tau)$ exists such that $|Q| = n$ and $\tau(\delta(q_0, (i)_k)) = a_i$ for all $i \in \mathbb{N}$. So, if for some N we prove that no k-DFAO $(Q, \Sigma_k, \delta, q_0, \Gamma, \tau)$ exists such that

$$|Q| = n \text{ and } \tau(\delta(q_0, (i)_k)) = a_i \text{ for all } 0 \leq i < N \qquad (*)$$

then we may conclude that $\|a\|_k > n$.

For a fixed number N and known values $a_0, a_1, \ldots, a_{N-1}$, the property $(*)$ is easily expressed in a formula in SMT format, and if a corresponding solver like Z3 [5] proves that it is unsatisfiable, then $\|a\|_k > n$ may be concluded. Proving $\|a\|_k^R > n$ is done similarly with the only difference that the k-ary representation of i is reversed.

For instance, we may prove $\|\mathsf{paper}\|_2 > 3$ for $\mathsf{paper} = 001001100011011\cdots$ as follows. Choose $N = 11$ and declare $d0, d1 : \{1,2,3\} \to \{1,2,3\}$ and $\tau : \{1,2,3\} \to \{0,1\}$. Here we identify $q_0 \in Q = \{1,2,3\}$ with $q_0 = 1$, and use $d0$ for $\delta(\cdot, 0)$ and $d1$ for $\delta(\cdot, 1)$. Now the requirements $\tau(\delta(q_0, (i)_k)) = \mathsf{paper}_i$ for $i = 0, 1, \ldots, 10$ are described by the following formula

$$\begin{aligned}
&\tau(1) = 0 \wedge \tau(d1(1)) = 0 \wedge \tau(d1(d0(1))) = 1 \wedge \tau(d1(d1(1))) = 0 \wedge \\
&\tau(d1(d0(d0(1)))) = 0 \wedge \tau(d1(d0(d1(1)))) = 1 \wedge \tau(d1(d1(d0(1)))) = 1 \wedge \\
&\tau(d1(d1(d1(1)))) = 0 \wedge \tau(d1(d0(d0(d0(1))))) = 0 \wedge \\
&\tau(d1(d0(d0(d1(1))))) = 0 \wedge \tau(d1(d0(d1(d0(1))))) = 1.
\end{aligned}$$

Putting this formula in SMT syntax and applying the SMT solver Z3 yields that this formula is unsatisfiable, so indeed $\|\mathsf{paper}\|_k > 3$ may be concluded. By choosing $N = 10$, so omitting the requirement $\tau(d1(d0(d1(d0(1))))) = 1$, then the formula is satisfiable, so $N = 11$ is the smallest value that does the job. With $N = 11$ and $n = 4$ the formula is satisfiable, and the satisfying assignment exactly describes the DFAO for paper we presented.

Current SMT solvers like Z3 easily deal with large values: building and solving formulas as above for $N \approx 10^4$ is no problem at all. For $n < 10$ typically satisfiability or unsatisfiability is established in a fraction of a second, and with slightly more computation time the approach is still feasible for $n = 10, 11, 12$. All claims on values of $\|a\|_k$ and $\|a\|_k^R$ for particular non-periodic sequences a in this paper have been checked in this way by Z3. This is done by a tool that starts by building the formula as above for $n = 2$ and then applies Z3, and as long as unsatisfiability is reported, the formula is built for $n = 3, 4, 5, \ldots$, until the smallest value n is found for which the formula is satisfiable, that is, describes a DFAO satisfying the requirement for the first N elements of a.

An alternative approach to prove that the property $(*)$ is unsatisfiable for the binary case is using Angluin's algorithm [2]. Then τ expresses whether a state is accepting or not. Then using an oracle stating whether a string is accepted or not, the minimal DFA is constructed for which for all strings from some finite set $(S \cup S.\Sigma).E$ it is known whether they are accepted or not. Then for $n <$ the size of this DFA and $N = 2^k$ for k being the length of the longest string in $(S \cup S.\Sigma).E$, the property $(*)$ is unsatisfiable. Exploiting this approach is a topic of further research.

It is a natural question for which number N the requirement for a_i has to hold for all $i < N$ such that we can safely conclude that it holds for all $i \in \mathbb{N}$. The key notion to deal with this question is equivalence of states in a DFAO with respect to a language. In a DFAO $(Q, \Sigma, \delta, q_0, \Gamma, \tau)$ two states $q, q' \in Q$ are called equivalent with respect to a language $L \subseteq \Sigma^*$, denoted by $q \equiv^L q'$ if $\tau(\delta(q,u)) = \tau(\delta(q',u))$ for all $u \in L$. For $n \in \mathbb{N}$ two states $q, q' \in Q$ are called n-equivalent with respect to L, denoted by $q \equiv^L_n q'$ if $\tau(\delta(q,u)) = \tau(\delta(q',u))$ for all $u \in L$ with $|u| \leq n$. It is clear that $\equiv^L_i$ is an equivalence relation for every $i \in \mathbb{N}$, and $q \equiv^L_{i+1} q'$ implies $q \equiv^L_i q'$ for every $q, q' \in Q$, $i \in \mathbb{N}$. For $L = \Sigma^*$ we often omit L in the notation $\equiv^L$ and $\equiv^L_i$.

Part (a) of the following lemma is an improvement of [8], Corollary 3.1 (page 59), stating that $\equiv$ and $\equiv_{n-1}$ coincide for $n = |Q|$ and $L = \Sigma^*$. For DFAs only it appears as Proposition 6.1 of Chapter III in [6]. We need it to prove parts (b) and (c) in which L consists of strings occurring as $(i)_k$, $(i)^R_k$, respectively.

Lemma 7. *Let $n = |Q|$.*

(a) *If $L = \Sigma^*$ then $\equiv^L$ and $\equiv^L_{n-2}$ coincide.*
(b) *If $\Sigma' \subseteq \Sigma$ and $L = \{\epsilon\} \cup \{xu \mid x \in \Sigma' \wedge u \in \Sigma^*\}$ then $\equiv^L$ and $\equiv^L_{n-1}$ coincide.*
(c) *If $\Sigma' \subseteq \Sigma$ and $L = \{\epsilon\} \cup \{ux \mid x \in \Sigma' \wedge u \in \Sigma^*\}$ then $\equiv^L$ and $\equiv^L_{n-1}$ coincide.*

Proof. First we prove part (a), so $L = \Sigma^*$, and we omit L in the notation. If $\equiv_0$ has only one equivalence class, then $\tau(q) = \tau(q')$ for all $q, q' \in Q$, hence $q \equiv q'$ for all $q, q' \in Q$, for which part (a) trivially holds. So $\equiv_0$ has at least two equivalence classes.

Next assume that $\equiv_i$ and $\equiv_{i+1}$ coincide for some $i \in \mathbb{N}$, and $q \equiv_{i+1} q'$. Exploiting $\delta(q, xu) = \delta(\delta(q,x),u)$ and $|xu| = |u|+1$ for $x \in \Sigma, u \in \Sigma^*$ we obtain

$$\begin{aligned} q \equiv_{i+2} q' &\iff q \equiv_{i+1} q' \wedge \forall x \in \Sigma : \delta(q,x) \equiv_{i+1} \delta(q',x) \\ &\iff q \equiv_i q' \wedge \forall x \in \Sigma : \delta(q,x) \equiv_i \delta(q',x) \\ &\iff q \equiv_{i+1} q' \end{aligned}$$

for all $q, q' \in Q$. Hence also $\equiv_{i+1}$ and $\equiv_{i+2}$ coincide, so $\equiv_i$ and $\equiv_j$ coincide for all $j > i$, from which we conclude that $\equiv_i$ and $\equiv$ coincide.

From the above we conclude that for increasing i, either $\equiv_{i+1}$ has strictly more equivalence classes then $\equiv_i$, or $\equiv_i$ coincides with $\equiv$. As $\equiv_0$ has at least two equivalence classes and all equivalence relations including $\equiv$ have at most $|Q| = n$ equivalence classes, at most $n-2$ strict increases are possible, proving part (a) of the lemma.

Next we prove part (b). Assume that $q \equiv^L_{n-1} q'$; we have to prove $q \equiv^L q'$. From $q \equiv^L_{n-1} q'$ and the definition of L follows that for every $x \in \Sigma'$ and every $u \in \Sigma^*$ with $|u| \leq n-2$ we have $\tau(\delta(\delta(q,x),u)) = \tau(\delta(\delta(q',x),u))$. So $\delta(q,x) \equiv_{n-2} \delta(q',x)$. From part (a) we conclude $\delta(q,x) \equiv \delta(q',x)$. Since this holds for every $x \in \Sigma'$, by the definition of L we obtain $q \equiv^L q'$.

It remains to prove part (c). For every $x \in \Sigma'$ we define a modified DFAO obtained by replacing τ by τ_x defined by $\tau_x(q) = \tau(\delta(q,x))$; for the corresponding equivalence relations with respect to the language Σ^* we write $\equiv^x$ and $\equiv^x$. From the definitions we obtain

$$q \equiv^L_{n-1} q' \iff \forall x \in \Sigma' : q \equiv^x_{n-2} q';$$

now part (c) follows from part (a) applied to the modified DFAOs. $\square$

Lemma 7, part (a), can not be improved on, in the sense that $\equiv_{n-3}$ and $\equiv$ may not coincide, for every $n \geq 3$, as is shown by the following example of a DFAO consisting of a cycle of length n. Take $Q = \{1, 2, \ldots, n\}$, $\Sigma = \{0\}$, $\delta(i, 0) = i+1$ for $i < n$, $\delta(n, 0) = 1$, $\tau(n) = 1$, $\tau(i) = 0$ for $i < n$. Then

- $\equiv_0$ has two equivalence classes $\{n\}$, $\{1, 2, \ldots, n-1\}$,
- $\equiv_1$ has three equivalence classes $\{n\}$, $\{n-1\}$, $\{1, 2, \ldots, n-2\}$,
- $\cdots$
- $\equiv_{n-3}$ has $n-1$ equivalence classes $\{n\}, \{n-1\}, \ldots, \{3\}, \{1, 2\}$,
- $\equiv$ coincides with $\equiv_{n-2}$ and has n equivalence classes, all being singletons.

Also Lemma 7, parts (b) and (c) are sharp: in the above example, extend Σ to $\{0, 1\}$ and define $\delta(q, 1) = q$ for all $q \in Q$. Then for $\Sigma' = \{1\}$ the equivalence relations $\equiv_0$ and $\equiv_1$ coincide, and $\equiv_{n-2}$ has $n-1$ equivalence classes $\{n\}, \{n-1\}, \ldots, \{3\}, \{1, 2\}$, and does not coincide with $\equiv$ having n equivalence classes all being singletons.

Theorem 8. *Let $M = (Q, \Sigma_k, \delta, q_0, \Gamma, \tau)$ be a k-DFAO and a be a k-automatic sequence.*

If $\tau(\delta(q_0,(i)_k)) = a_i$ *for* $i < k^{\|a\|_k+|Q|-1}$*, then* $\tau(\delta(q_0,(i)_k)) = a_i$ *for all* $i \in \mathbb{N}$*.*
If $\tau(\delta(q_0,(i)_k^R)) = a_i$ *for* $i < k^{\|a\|_k^R+|Q|-1}$*, then* $\tau(\delta(q_0,(i)_k^R)) = a_i$ *for all* $i \in \mathbb{N}$*.*

Proof. For the first claim let $M' = (Q', \Sigma_k, \delta', q'_0, \Gamma, \tau')$ such that $|Q'| = \|a\|_k$ and $\tau'(\delta'(q'_0,(i)_k)) = a_i$ for all $i \in \mathbb{N}$. Now we take the disjoint union of M and M' of size $n = \|a\|_k + |Q|$. Let $\Sigma' = \Sigma_k \setminus \{0\}$ and $L = \{\epsilon\} \cup \{xu \mid x \in \Sigma' \wedge u \in \Sigma_k^*\}$. Observe that the strings $(i)_k$ for $i < k^{\|a\|_k+|Q|-1}$ exactly coincide with the strings in L of length $\leq n-1$. So from the assumption we conclude $q_0 \equiv_{n-1}^L q'_0$. Now Lemma 7, part (b), yields $q_0 \equiv^L q'_0$, proving the theorem.

The proof of the second claim is similar, using Lemma 7, part (c). □

For both claims in Theorem 8 the bound is sharp in the sense that the condition can not be weakened to $\tau(\delta(q_0,(i)_k)) = a_i$ for all $i < k^{\|a\|_k+|Q|-2}$. This is shown by the following example. Define a by $a_i = 1$ if the number of 0s in $(i)_k$ is $n-2$ modulo n, and $a_i = 0$ otherwise. Then $\|a\|_k = \|a\|_k^R = n$. Define $Q = \{1,2,\ldots,n-1\}$, $\delta(q,0) = q+1$ for $q < n-1$, $\delta(n-1,0) = 1$, $\delta(q,j) = q$ for all $q \in Q$, $0 < j < k$, $\tau(n-2) = 1$, $\tau(q) = 0$ for $i \neq n-2$. Then $\tau(\delta(q_0,(i)_k)) = a_i$ for all i for which $(i)_k$ or $(i)_k^R$ contains fewer than $2n-3 = \|a\|_k + |Q| - 2$ zeros, and every $i < k^{\|a\|_k+|Q|-2}$ has strictly fewer than $2n-3 = \|a\|_k + |Q| - 2$ zeros.

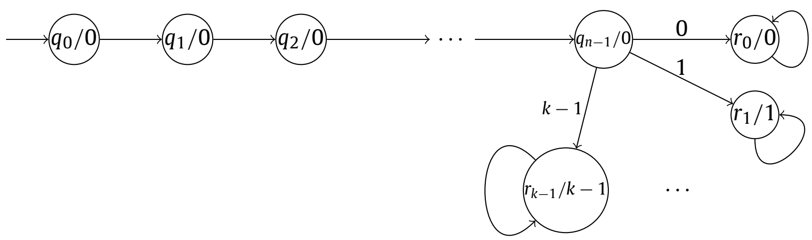
For a concrete sequence a for which we know $\|a\|_k \leq m$ or $\|a\|_k^R \leq m$, we may conclude that $\|a\|_k = n$ or $\|a\|_k^R = n$ by Theorem 8 by showing satisfiability of the formula expressing $\tau(\delta(q_0,(i)_k)) = a_i$ or $\tau(\delta(q_0,(i)_k^R)) = a_i$ in a DFAO of n states for all $i < N$ for $N = k^{n+m-1}$. This is the default way we follow for claims of the shape $\|a\|_k = n$ or $\|a\|_k^R = n$ for particular sequences a throughout the paper.

4. The exponential gap

The following theorem shows that there can be an exponential gap between $\|a\|_k$ and $\|a\|_k^R$, in both directions. Its proof is inspired by the folklore result that the language $(0+1)^*1(0+1)^{n-1}$ has an NFA of size $n+1$, and its reverse has a DFA of size $n+1$, but its smallest DFA has size at least 2^n. We found it in [10], Section 3.2, page 67, exercise 3. Many similar results on state complexity are known, e.g., in [9] it is proved that all values up to 2^n can be reached as sizes.

Theorem 9. *For every* $n > 1$ *there exist* k*-automatic sequences* a, b *such that* $\|a\|_k \leq n+k$ *and* $\|a\|_k^R \geq (k-1)k^{n-1}$*, and* $\|b\|_k^R \leq n+k$ *and* $\|b\|_k \geq (k-1)k^{n-1}$*.*

Proof. Define a by $a_i = 0$ for $i < k^n$, and $a_i = j$ if and only if the nth digit of $(i)_k$ is j, for $j = 0,1,\ldots,k-1$, $i \geq k^n$. The following DFAO satisfies $\tau_M(\delta_M(q_0,(i)_k)) = a_i$ by construction:



Here all unlabeled arrows are assumed to be labeled by all symbols $0,1,\ldots,k-1$. Since this DFAO has $n+k$ states we obtain $\|a\|_k \leq n+k$.

To prove $\|a\|_k^R \geq (k-1)k^{n-1}$ we apply Lemma 4. For $i = 1,2,\ldots,(k-1)k^{n-1}$ define $m_i = k^n + i - 1$, so the numbers m_i are exactly the numbers of k-ary length n, starting in a digit $\neq 0$. For any two distinct such numbers m_i and m_j there is a position p on which they differ, so by choosing $v = 1^{n-p}$, the strings $v(m_i)_k$ and $v(m_j)_k$ differ in their n-th position. So the condition of Lemma 4 holds and we conclude $\|a\|_k^R \geq (k-1)k^{n-1}$.

Define b by $b_i = 0$ for $i < k^n$, and $a_i = j$ if and only if the nth element of $(i)_k^R$ is j, for $j = 0,1,\ldots,k-1$, $i \geq k^n$. A similar argument using the same automaton proves the claim for b. □

5. The k-kernel

For $j \in \Sigma_k$ we define $p_j(a) = a_j a_{k+j} a_{2k+j} a_{3k+j} \cdots$ by $(p_j(a))_i = a_{ik+j}$ for all $i \in \mathbb{N}$. So for $k = 2$ we have $p_0(a) = \mathsf{even}(a) = a_0 a_2 a_4 \cdots$ and $p_1(a) = \mathsf{odd}(a) = a_1 a_3 a_5 \cdots$.

For an infinite sequence $a = a_0 a_1 a_2 a_3 \cdots$ over Γ we define its *k-kernel* $K_k(a)$ to be the smallest set $K_k(a) \subseteq \Gamma^{\mathbb{N}}$ such that

- $a \in K_k(a)$,
- for every $b \in K_k(a)$ and every $j \in \Sigma_k$ we have $p_j(b) \in K_k(a)$.

We recall from [6], Prop. V.3.3, or [1], Theorem 6.6.2, that a is k-automatic if and only if $K_k(a)$ is finite.

For a k-automatic sequence $a = a_0a_1a_2a_3\cdots$ over the alphabet Γ its *k-kernel* $K_k(a)$ has a natural DFAO structure: the DFAO $\mathcal{K}_k(a) = (K_k(a), \Sigma_k, \delta, a, \Gamma, \tau)$, where

- the input alphabet is Σ_k,
- $K_k(a)$ is the set of states,
- $\delta : K_k(a) \times \Sigma_k \to Q$ is defined by $\delta(q, j) = p_j(q)$,
- a is the initial state,
- the output alphabet is Γ,
- the output function $\tau : K_k(a) \to \Sigma_k$ is defined by $\tau(b_0b_1b_2\cdots) = b_0$.

Recall that for $k = 2$ we have $p_0 = \mathsf{even}$ and $p_1 = \mathsf{odd}$, so in $K_2(a)$ the 0-steps describe even and the 1-steps describe odd. For thue (Example 2) the 2-kernel exactly coincides with the DFAO M_{thue} given in Section 2, in which q_0 coincides with thue and q_1 coincides with the sequence obtained from thue by swapping symbols 0 and 1. For paper (Example 3) the 2-kernel exactly coincides with the given DFAO M_{paper^R}, in which q_0 coincides with paper, q_1 with $(01)^\omega = 010101\cdots$, q_2 with $0^\omega = 000\cdots$ and q_3 with $1^\omega = 111\cdots$.

The following theorem is straightforwardly proved by induction on i:

Theorem 10. *For every k-automatic sequence $a = a_0a_1a_2a_3\cdots$ and every $i \in \mathbb{N}$ we have $\tau(\delta(a, (i)_k^R)) = a_i$ where τ, δ refer to $\mathcal{K}_k(a) = (K_k(a), \Sigma_k, \delta, a, \Gamma, \tau)$.*

As a consequence, by only giving the DFAO $\mathcal{K}_k(a)$ the sequence a is fully defined.

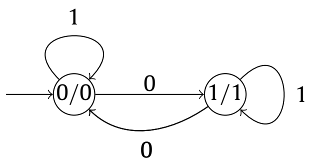
Theorem 11. *Up to isomorphism, the DFAO $\mathcal{K}_k(a)$ is the unique DFAO of minimal size such that $\tau(\delta(a, (i)_k^R 0^j)) = a_i$ for every $i, j \in \mathbb{N}$.*

Proof. Let $\mathcal{K}_k(a) = (K_k(a), \Sigma_k, \delta, a, \Gamma, \tau)$. Combining Theorem 10 with the fact that $\tau(q) = \tau(\delta(q, 0))$ for all $q \in K_k(a)$ yields $\tau(\delta(a, (i)_k^R 0^j)) = a_i$ for every $i, j \in \mathbb{N}$. Assume it is not of minimal size with this property. Then there are two distinct states q, q' such that $\tau(\delta(q, u)) = \tau(\delta(q', u))$ for all $u \in \Sigma_k^*$. Since q, q' are sequences over Σ_k, applying Theorem 10 to $\mathcal{K}_k(q)$ and $\mathcal{K}_k(q')$ yield $q_i = q'_i$ for all $i \in \mathbb{N}$. But then q, q' are equal as sequences, contradicting that they are distinct. □

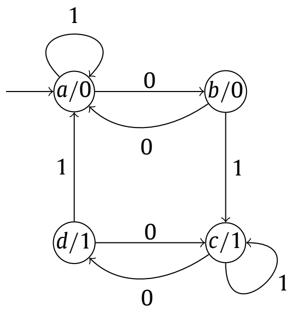
Recall that $\|a\|_k^R$ is the minimal size $|Q|$ for which a DFAO $M = (Q, \Sigma, \delta, q_0, \Gamma, \tau)$ exists such that $\tau(\delta(q_0, (i)_k^R)) = a_i$ for every $i \in \mathbb{N}$. We observe that a DFAO with this property does not need to be unique. For instance, for $a = 01^\omega$ the DFAO $\mathcal{K}_k(a)$ is a minimal DFAO with this property, having two states a and $b = 1^\omega$, and $\delta(a, 0) = a, \delta(a, 1) = \delta(b, 0) = \delta(b, 1) = b$, $\tau(a) = 0$, $\tau(b) = 1$. But the DFAO with the same two states a, b and $\delta(b, 0) = a, \delta(a, 0) = \delta(a, 1) = \delta(b, 1) = b$, $\tau(a) = 0$, $\tau(b) = 1$ produces the same sequence $a = 01^\omega$.

In the following example we observe that $\|a\|_k^R$ can be strictly smaller than $|K_k(a)|$, the size of the state space of $\mathcal{K}_k(a)$.

Example 12. Define $a_i = 1$ if the number of zeros in $(i)_2$ is odd, and $a_i = 0$ if this number is even. Clearly it admits the following DFAO, in which as usual $\tau(q) = x$ is denoted by q/x in the state q:



Hence $\|a\|_k^R \leq 2$; we obtain $\|a\|_k^R = 2$ since the sequence contains both 0 and 1. However, $|K_k(a)| = 4$, since $\mathcal{K}_k(a)$ is the following DFAO:



The sequences a, b, c, d are as follows:

$$a = 001001101\cdots, \; b = 010110010\cdots,$$
$$c = 110110010\cdots, \; d = 101001101\cdots.$$

Observe that a and d differ only at the first position, and similarly for b and c. The next lemma states that this always occurs if $|K_k(a)|$ is greater than $\|a\|_k^R$.

Lemma 13. *Let $\mathcal{K}_k(a) = (K_k(a), \Sigma_k, \delta, a, \Gamma, \tau)$ be the kernel of sequence a over Γ and let $(Q_M, \Sigma_k, \delta_M, q_0, \Gamma, \tau_M)$ be such that $\tau_M(\delta_M(q_0, (i)_k^R)) = a_i$ for all $i \in \mathbb{N}$. Assume that $\delta_M(q_0, u) = \delta_M(q_0, v)$ for $u, v \in \Sigma_k^*$. Then*

$$\delta(a, u)_i = \delta(a, v)_i \text{ for all } i > 0.$$

Proof. Let $i > 0$. For any $w \in \Sigma_k^*$ define the numbers m_w by $(m_w)_k = (i)_k w^R$; this is possible since $(i)_k w^R$ does not start in 0 since $i > 0$. For any $b \in K_k(a)$ we obtain $b_i = \tau(\delta(b, (i)_k^R))$ by considering $\mathcal{K}_k(b)$. Hence

$$\delta(a, w)_i = \tau(\delta(\delta(a, w), (i)_k^R)) = \tau(\delta(a, w(i)_k^R)) = \tau(\delta(a, (m_w)_k^R)) = a_{m_w}.$$

We obtain:
$$\begin{aligned} \delta(a, u)_i = a_{m_u} &= \tau_M(\delta_M(q_0, (m_u)_k^R)) \\ &= \tau_M(\delta_M(q_0, u(i)_k^R)) \\ &= \tau_M(\delta_M(\delta_M(q_0, u), (i)_k^R)) \\ &= \tau_M(\delta_M(\delta_M(q_0, v), (i)_k^R)) \\ &= \tau_M(\delta_M(q_0, (m_v)_k^R)) = a_{m_v} = \delta(a, v)_i. \quad \square \end{aligned}$$

Theorem 14. *Let a be a k-automatic sequence over an alphabet Γ. Then*

$$\|a\|_k^R \le |K_k(a)| \le |\Gamma| \cdot \|a\|_k^R.$$

Moreover, if a is periodic then $\|a\|_k^R = |K_k(a)|$.

Proof. The inequality $\|a\|_k^R \le |K_k(a)|$ holds since the automaton $\mathcal{K}_k(a)$ satisfies $\tau(\delta(a, (i)_k^R)) = a_i$ for every $i \in \mathbb{N}$. For the other inequality let $M = (Q, \Sigma, \delta, q_0, \Gamma, \tau)$ be a DFAO of minimal size $\|a\|_k^R$ such that $\tau(\delta(q_0, (i)_k^R)) = a_i$ for every $i \in \mathbb{N}$. For every $b \in K_k(a)$ choose $u_b \in \Sigma_K^*$ such that $b = \delta(a, u_b)$. Define $\sim$ on $K_k(a)$ by $b \sim c \iff \delta_M(q_0, u_b) = \delta_M(q_0, u_c)$.

According to Lemma 13 $b \sim c$ implies that $b_i = c_i$ for all $i > 0$, so the difference between b and c may only be caused by $b_0 \neq c_0$. Hence every equivalence class of $\sim$ has at most $|\Gamma|$ elements, while the number of equivalence classes is $|Q| = \|a\|_k^R$. This proves $|K_k(a)| \le |\Gamma| \cdot \|a\|_k^R$.

In case a is periodic then all elements of $K_k(a)$ are periodic too, and $b_i = c_i$ for all $i > 0$ implies $b = c$. Hence in that case all equivalence classes consist of a single element, proving $\|a\|_k^R = |K_k(a)|$. $\square$

6. Morphic sequences

Recall that $\|a\|_k = |Q_M|$ for the smallest Q_M being the set of states of a DFAO $M = (Q_M, \Sigma_k, \delta_M, q_0, \Gamma, \tau_M)$ for which $\tau_M(\delta_M(q_0, (i)_k)) = a_i$ for every $i \in \mathbb{N}$. Again this DFAO of minimal size is not unique: for $a = 01^\omega$ the DFAO $\mathcal{K}_k(a)$ as given above also satisfies $\tau_M(\delta_M(q_0, (i)_k)) = a_i$ for all $i \in \mathbb{N}$, but after changing $\delta(a, 0) = a$ to $\delta(a, 0) = b$ this property still holds, since $(i)_k$ never starts by 0.

Just like $\|a\|_k^R$ is strongly related to the kernel of a as described in Theorem 14, $\|a\|_k$ is strongly related to the number of symbols needed to describe a as a morphic sequence with respect to a k-uniform morphism. Morphic sequences are defined as follows. For a morphism $h : \Delta \to \Delta^*$ and some $x \in \Delta$ satisfying $h(x) = xu$ we observe that the string $h^i(x)$ equals $xuh(u)h^2(u)\cdots h^{i-1}(u)$ for every $i > 0$, so yielding a limit $h^\omega(x) = xuh(u)h^2(u)h^3(u)\cdots$, typically being an infinite sequence which is a fixed point of h.

A sequence a over an alphabet Γ is called *morphic* with respect to a morphism $h : \Delta \to \Delta^*$ and a coding $\tau : \Delta \to \Gamma$ if $a = \tau(h^\omega(x))$ for some $x \in \Delta$ satisfying $h(x) = xu$. The morphism $h : \Delta \to \Delta^*$ is called k-uniform if the string $h(y) \in \Delta^*$ has length k for every $y \in \Delta$. It is well-known (Cobham [4], see also [1] Theorem 6.3.2) that a is k-automatic if and only if it is morphic with respect to a k-uniform morphism. For instance, for the examples from Example 2 and Example 3 we have $\mathsf{thue} = h^\omega(0)$ for $h(0) = 01, h(1) = 10$, and $\mathsf{paper} = \tau(g^\omega(0))$ for $g(0) = 02$, $g(1) = 31$, $g(2) = 32$, $g(3) = 01$, $\tau(0) = \tau(2) = 0$, $\tau(1) = \tau(3) = 1$.

Theorem 15. *Let a be a k-automatic sequence. Let $d(a)$ be the minimal size of the alphabet Δ such that $a = \tau(h^\omega(x))$ for a k-uniform morphism $h : \Delta \to \Delta^*$ and a coding $\tau : \Delta \to \Gamma$. Then $\|a\|_k \le d(a) \le \|a\|_k + 1$.*

Proof. The k-DFAO $M = (\Delta, \Sigma_k, \delta, q_0, \Gamma, \tau)$ with $q_0 = x$ and $\delta(q, y) = h(q)_y$, where we write $h(q) = h(q)_0 \cdots h(q)_{k-1}$, satisfies $\tau(\delta(q_0, (i)_k)) = a_i$ for all $i \ge 0$ as is showed in the proof of Theorem 6.3.2 of [1]. As $\|a\|_k$ is the smallest size of a k-DFAO with this property we obtain $\|a\|_k \le d(a)$.

Conversely, if $M = (Q_M, \Sigma_k, \delta_M, q_0, \Gamma, \tau_M)$ is a k-DFAO of size $\|a\|_k$ with $\tau_M(\delta_M(q_0, (i)_k)) = a_i$ for all $i \ge 0$, then by choosing a fresh state q'_0 and defining $Q = Q_M \cup \{q'_0\}$, $\delta(q, y) = \delta_M(q, y)$ for $q \in Q_M$, $\delta(q'_0, 0) = q'_0$, $\delta(q'_0, y) = \delta_M(q_0, y)$ for $y \ne 0$, $\tau(q'_0) = \tau_M(q_0)$, $\tau(q) = \tau_M(q)$ for $q \in Q_M$, we obtain the k-DFAO $(Q, \Sigma_k, \delta, q'_0, \Gamma, \tau)$ of size $\|a\|_k + 1$ with $\tau(\delta(q'_0, (i)_k)) = a_i$ for all $i \ge 0$. Using the fact that $\delta(q'_0, 0) = q'_0$ we obtain $a = \tau(h^\omega(q'_0))$ for h defined by $h(q) = \delta(q, 0)\delta(q, 1) \cdots \delta(q, k-1)$ as is shown in the proof of Theorem 6.3.2 of [1]. Hence $d(a) \le \|a\|_k + 1$. □

7. The effect of basic operations

For any sequence $a = a_0 a_1 a_2 a_3 \cdots$ we define its tail $\mathsf{tail}(a) = a_1 a_2 a_3 a_4 \cdots$ by $(\mathsf{tail}(a))_i = a_{i+1}$ for all $i \in \mathbb{N}$.

Theorem 16. *For any k-automatic sequence a we have $\|\mathsf{tail}(a)\|^R_k \le 2\|a\|^R_k$ and $\|\mathsf{tail}(a)\|_k \le (\|a\|_k)^2$. For every $n > 1$ there exists a k-automatic sequence a such that $\|a\|_k = n$ and $\|\mathsf{tail}(a)\|_k = n^2$.*

Proof. For the first claim take a DFAO $M = (Q, \Sigma_k, \delta, q_0, \Gamma, \tau)$ of size $\|a\|^R_k$ with $\tau(\delta(q_0, (i)^R_k)) = a_i$ for all $i \ge 0$. Let m be the smallest number $m > 0$ such that there exists $j < m$ such that $\delta(q_0, 0^m) = \delta(q_0, 0^j)$. Since there are $\|a\|^R_k$ states we obtain $m \le \|a\|^R_k$. Introduce fresh states $r_0, \ldots, r_{m-1}$ and define the DFAO $M' = (Q \cup \{r_0, \ldots, r_{m-1}\}, \Sigma_k, \delta', r_0, \Gamma, \tau')$ by

$$\delta'(q, x) = \delta(q, x) \text{ for } q \in Q, x \in \Sigma_k,$$

$$\delta'(r_i, k-1) = r_{i+1} \text{ for } i = 1, \ldots, m-2,$$

$$\delta'(r_{m-1}, k-1) = r_j \text{ for } j < m \text{ with } \delta(q_0, 0^m) = \delta(q_0, 0^j),$$

$$\delta'(r_i, x) = \delta(q_0, 0^i(x+1)) \text{ for } i = 0, \ldots, m-1, x < k-1.$$

By construction we have $\delta'(r_0, (k-1)^i x) = \delta(q_0, 0^i(x+1))$ for all $i \in \mathbb{N}$, $x < k-1$. So by defining $\tau'(q) = \tau(q)$ for $q \in Q$ and $\tau'(r_i) = \tau(\delta(q_0, 0^i))$ for $i = 0, \ldots, m-1$ we obtain

$$\tau'(\delta'(r_0, (vx(k-1)^i)^R)) = \tau(\delta(q_0, (v(x+1)0^i)^R))$$

and

$$\tau'(\delta'(r_0, (k-1)^i)) = \tau(\delta(q_0, (10^i)^R))$$

for all $i \in \mathbb{N}$, $v \in \Sigma^*_k$. Since $[vx(k-1)^i]_k + 1 = [v(x+1)0^i]_k$, and $[(k-1)^i]_k + 1 = [10^i]_k$, and every number in $\mathbb{N}$ is either of the shape $[vx(k-1)^i]_k$ or $[(k-1)^i]_k$, this proves that M' is a DFAO for $\mathsf{tail}(a)$. Since $|Q \cup \{r_0, \ldots, r_{m-1}\}| \le 2|Q|$ this yields $\|\mathsf{tail}(a)\|^R_k \le 2\|a\|^R_k$.

For the second claim take a DFAO $M = (Q, \Sigma_k, \delta, q_0, \Gamma, \tau)$ of size $\|a\|_k$ with $\tau(\delta(q_0, (i)_k)) = a_i$ for all $i \ge 0$. Define the DFAO $\overline{M} = (Q \times Q, \Sigma_k, \overline{\delta}, \overline{q_0}, \Gamma, \overline{\tau})$ of size $(\|a\|_k)^2$ by

$$\overline{q_0} = (q_0, \delta(q_0, 1)), \quad \overline{\tau}(q, q') = \tau(q'),$$

$$\overline{\delta}((q, q'), k-1) = (\delta(q, k-1), \delta(q', 0)),$$

$$\overline{\delta}((q, q'), x) = (\delta(q, x), \delta(q, x+1)),$$

for all $q, q' \in Q$, $x < k-1$. For every $i \in \mathbb{N}$ we have either $(i)_k = (k-1)^m$ or $(i)_k = vx(k-1)^m$, for some $m \ge 0$, $v \in \Sigma^*_k$, $x < k-1$. In the first case we have $(i+1)_k = 10^m$, in the second case $(i+1)_k = v(x+1)0^m$. The DFAO $\overline{M}$ has been constructed in such a way that $\overline{\tau}(\overline{\delta}(\overline{q_0}, (k-1)^m)) = \tau(\delta(q_0, 10^m))$ and $\overline{\tau}(\overline{\delta}(\overline{q_0}, vx(k-1)^m)) = \tau(\delta(q_0, v(x+1)0^m))$. Hence for all $i \in \mathbb{N}$ we have $\overline{\tau}(\overline{\delta}(\overline{q_0}, (i)_k)) = \tau(\delta(q_0, (i+1)_k)) = a_{i+1} = \mathsf{tail}(a)_i$, proving the second claim.

As $\|\mathsf{tail}(a)\|_k \le n^2$, for the last claim it suffices to prove $\|\mathsf{tail}(a)\|_k \ge n^2$. We define a by $a_i = 1$ if the number of zeros in $(i)_k$ is divisible by n, and $a_i = 0$ otherwise. A DFAO consisting of a single n-cycle easily produces a, so $\|a\|_k \le n$, and since a smaller one is not possible we obtain $\|a\|_k = n$. Let $b = \mathsf{tail}(a)$, so $b_i = a_{i+1}$ for all $i \in \mathbb{N}$. We prove $\|\mathsf{tail}(a)\|_k \ge n^2$ by Lemma 4. Choose $m_1, m_2, \ldots, m_{n^2}$ to be the numbers $[10^p(k-1)^q]_k$ for $p, q = 1, \ldots, n$. Let $m_i = [10^p(k-1)^q]_k$ and $m_j = [10^{p'}(k-1)^{q'}]_k$ for $i \ne j$, then $(p, q) \ne (p', q')$.

First we consider the case where $p+q$ and $p'+q'$ are distinct modulo n, choose r such that $p+q+r-1$ is divisible by n and $p'+q'+r-1$ is not. Choose $v = (k-1)^r$. Then $b_{[(m_i)_k v]_k} = a_{[(m_i)_k v]_k + 1} = a_{[10^{p-1}10^{q+r}]_k} = 1 \ne 0 = a_{[10^{p'-1}10^{q'+r}]_k} = b_{[(m_j)_k v]_k}$.

In the remaining case $p+q$ and $p'+q'$ are equal modulo n, and since $(p,q) \neq (p',q')$ we obtain that p and p' are distinct modulo n. Choose r such that $p+r$ is divisible by n and $p'+r$ is not. Choose $v = 0^{r+1}$, then $b_{[(m_i)_k v]_k} = a_{[(m_i)_k v]_k+1} = a_{[10^p(k-1)^q 0^r 1]_k} = 1 \neq 0 = a_{[10^{p'}(k-1)^{q'} 0^r 1]_k} = b_{[(m_j)_k v]_k}$.

So the conditions of Lemma 4 hold, and $\|\mathrm{tail}(a)\|_k \geq n^2$. $\square$

For our examples thue and paper from Example 2, 3, we have $\|\mathrm{tail(thue)}\|_2 = 4$, $\|\mathrm{tail(thue)}\|_2^R = 3$, $\|\mathrm{tail(paper)}\|_2 = 8$ and $\|\mathrm{tail(paper)}\|_2^R = 6$.

For any sequence $a = a_0a_1a_2a_3\cdots$ over Γ, and $x \in \Gamma$ the sequence $x \cdot a = xa_0a_1a_2a_3\cdots$ is defined by $(x \cdot a)_0 = x$ and $(x \cdot a)_i = a_{i-1}$ for all $i > 0$. The next theorem states that the effect of this operator $x\cdot$ adding an initial element x is similar to the operator tail.

Theorem 17. *For any k-automatic sequence a over Γ, and $x \in \Gamma$ we have $\|x \cdot a\|_k^R \leq 2\|a\|_k^R$ and $\|x \cdot a\|_k \leq (\|a\|_k)^2$. For every $n > 1$ there exists a k-automatic sequence a such that $\|a\|_k = n$ and $\|x \cdot a\|_k \geq n^2$.*

Proof. Similar to the proof of Theorem 16, with the roles of the symbols 0 and $k-1$ swapped, exploiting the property $[vx0^i]_k - 1 = [v(x-1)(k-1)^i]_k$ for any string v and any $x > 0$. $\square$

For our examples thue and paper from Example 2, 3 we have $\|0 \cdot \mathrm{thue}\|_2 = 4$, $\|0 \cdot \mathrm{thue}\|_2^R = 4$, $\|0 \cdot \mathrm{paper}\|_2 = 4$ and $\|0 \cdot \mathrm{paper}\|_2^R = 4$.

Recall that for $j \in \Sigma_k$ the operator p_j on sequences a is defined by $(p_j(a))_i = a_{ik+j}$ for all $i \in \mathbb{N}$.

Theorem 18. *For any k-automatic sequence a and $j \in \Sigma_k$ we have $\|p_j(a)\|_k \leq \|a\|_k$ and $\|p_j(a)\|_k^R \leq \|a\|_k^R$.*

Proof. Let $M = (Q, \Sigma_k, \delta, q_0, \Gamma, \tau)$ be a DFAO of size $\|a\|_k$ with $\tau(\delta(q_0, (i)_k)) = a_i$ for all $i \geq 0$. Define $M' = (Q, \Sigma_k, \delta, q_0, \Gamma, \tau')$ by $\tau'(q) = \tau(\delta(q, j))$ for all $q \in Q$. Then

$$(p_j(a))_i = a_{ki+j} = \tau(\delta(q_0, (i)_k j)) = \tau(\delta(\delta(q_0, (i)_k), j)) = \tau'(\delta(q_0, (i)_k)$$

for all $i \in \mathbb{N}$, so M' is a DFAO of size $\|a\|_k$ producing $p_j(a)$, so $\|p_j(a)\|_k \leq \|a\|_k$.

For the other claim let $M = (Q, \Sigma_k, \delta, q_0, \Gamma, \tau)$ be a DFAO of size $\|a\|_k^R$ with $\tau(\delta(q_0, (i)_k^R)) = a_i$ for all $i \geq 0$. Define $M' = (Q, \Sigma_k, \delta, \delta(q_0, j), \Gamma, \tau)$. Then

$$(p_j(a))_i = a_{ki+j} = \tau(\delta(q_0, j(i)_k)^R) = \tau(\delta(\delta(q_0, j), (i)_k)^R)$$

for all $i \in \mathbb{N}$, so M' is a DFAO of size $\|a\|_k^R$ producing $p_j(a)$, so $\|p_j(a)\|_k \leq \|a\|_k$. $\square$

For our examples thue and paper from Example 2, 3 we have $\|\mathrm{even(thue)}\|_2 = 2$, $\|\mathrm{odd(thue)}\|_2^R = 2$, $\|\mathrm{even(paper)}\|_2 = 2$ and $\|\mathrm{odd(paper)}\|_2^R = 4$.

When applying an operator $f : \Gamma_1 \times \Gamma_2 \to \Gamma_3$ on two sequences $a \in \Gamma_1^{\mathbb{N}}$, $b \in \Gamma_2^{\mathbb{N}}$, by $f(a,b) \in \Gamma_3^{\mathbb{N}}$ we mean the sequence defined by $f(a,b)_i = f(a_i, b_i)$ for all $i \in \mathbb{N}$. For instance, $\wedge$ applied on boolean sequences denotes the elementwise conjunction of the two boolean sequences. Corollary 5.4.5 from [1] states that if a and b are k-automatic, then $f(a,b)$ is k-automatic.

Theorem 19. *For any two k-automatic sequences $a \in \Gamma_1^{\mathbb{N}}$, $b \in \Gamma_2^{\mathbb{N}}$ and every function $f : \Gamma_1 \times \Gamma_2 \to \Gamma_3$ we have $\|f(a,b)\|_k \leq \|a\|_k \cdot \|b\|_k$ and $\|f(a,b)\|_k^R \leq \|a\|_k^R \cdot \|b\|_k^R$.*

Proof. Let $(Q_1, \Sigma_k, \delta_1, q_{10}, \Gamma_1, \tau_1)$ be a DFAO of size $\|a\|_k$ with $\tau_1(\delta_1(q_{10}, (i)_k)) = a_i$ for all $i \geq 0$. Let $(Q_2, \Sigma_k, \delta_2, q_{20}, \Gamma_2, \tau_2)$ be a DFAO of size $\|b\|_k$ with $\tau_2(\delta_2(q_{20}, (i)_k)) = b_i$ for all $i \geq 0$. Then $(Q_1 \times Q_2, \Sigma_k, \delta, (q_{10}, q_{20}), \Gamma_3, \tau)$ for δ, τ defined by $\delta((q_1, q_2), x) = (\delta_1(q_1, x), \delta_2(q_2, x))$ and $\tau(q_1, q_2) = f(\tau_1(q_1), \tau_2(q_2))$ for all $q_1 \in Q_1, q_2 \in Q_2, x \in \Sigma_k$, is a DFAO of size $\|a\|_k \cdot \|b\|_k$ for $f(a,b)$. The proof for the reversed version is similar. $\square$

Combining our examples thue and paper from Example 2, 3 we have $\|\mathrm{thue} \wedge \mathrm{paper}\|_2 = 8$ and $\|\mathrm{thue} \wedge \mathrm{paper}\|_2^R = 7$. This is proved by our SMT approach. The latter can be proved in two ways. The first is by Theorem 8 for which we need to consider the first $2^{14} = 16384$ elements of the sequence. The second way is by observing that the DFAO of size 8 from the construction of Theorem 19 has two equivalent states that can be shared yielding a DFAO of 7 states, while for the SMT approach much less elements are needed to conclude that it can not decrease further.

A next issue is taking *arithmetic subsequences*. For $n > 0, m \geq 0$ and any sequence a we define the sequence $a[m,n]$ by $a[m,n]_i = a_{in+m}$ for all $i \geq 0$. Theorem 6.8.1 from [1] states that if a is k-automatic, then so is $a[m,n]$. So it is a natural question what is the complexity of such an arithmetic subsequence. The next theorem gives an answer for $\|\cdot\|_k^R$.

Theorem 20. *Let a be a k-automatic sequence and* $0 \le m < n$. *Then*

$$\|a[n,m]\|_k^R \le n \cdot \|a\|_k^R.$$

Proof. Let $M = (Q, \Sigma_k, \delta, q_0, \Gamma, \tau)$ be a DFAO of size $\|a\|_k^R$ with $\tau(\delta(q_0, (i)_k^R)) = a_i$ for all $i \ge 0$. The goal is to construct a DFAO M' of size $n|Q|$ representing $a[n,m]$.

For $q \in Q$ we define the sequence a^q by $a_i^q = \tau(\delta(q, (i)_k^R))$ for all $i \ge 0$. We define

$$M' = (Q \times \{0, 1, \dots, n-1\}, \Sigma_k, \delta', (q_0, m), \Gamma, \tau')$$

for δ', τ' defined by

$$\delta'((q,x), j) = (\delta(q, (jn+x) \bmod k), (jn+x) \div k),$$
$$\tau'(q,x) = a_x^q,$$

for all $0 \le j < k$, $0 \le x < n$, where as usual mod, $\div$ are defined by $y = k(y \div k) + (y \bmod k)$ and $0 \le (y \bmod k) < k$, for all $y \ge 0$. Note that δ' is well-defined since $0 \le jn + x < kn$, so $0 \le (jn+x) \div k < n$.

Now the theorem follows from the following property

$$a_{in+m} = \tau'(\delta'((q_0, m), (i)_k^R))$$

for all $i \ge 0$, which is proved by induction on i of the more general property

$$a_{in+x}^q = \tau'(\delta'((q,x), (i)_k^R))$$

for all $0 \le x < n$, $q \in Q$. For $i = 0$ we have

$$a_x^q = \tau'(q,x) = \tau'(\delta'((q,x), \epsilon)) = \tau'(\delta'((q,x), (0)_k^R)).$$

For $i > 0$ we write $in + x = (i'n + x')k + y$ for $i'n + x = (in+x) \div k$ and $y = (in+x) \bmod k$ and $0 \le x' < n$. So $i' = i \div k$ and $x' = ((i \bmod k)n + x) \div k$.

Since $0 \le i' < i$ we may apply the induction hypothesis on i'. We obtain

$$\begin{aligned}
a_{in+x}^q &= \tau(\delta(q, (in+x)_k^R)) \\
&= \tau(\delta(q, ((i'n+x')k + y)_k^R)) \\
&= \tau(\delta(q, y(i'n+x')_k^R)) \\
&= \tau(\delta(\delta(q,y), (i'n+x')_k^R)) \\
&= a_{i'n+x'}^{\delta(q,y)} \qquad \text{(IH on } i'\text{)} \\
&= \tau'(\delta'((\delta(q,y), x'), (i')_k^R)) \\
&= \tau'(\delta'(\delta(q, ((i \bmod k)n + x) \bmod k, ((i \bmod k)n + x) \div k), (i')_k^R)) \\
&= \tau'(\delta'(\delta'((q,x), (i \bmod k)), (i')_k^R)) \\
&= \tau'(\delta'((q,x), (i \bmod k)(i')_k^R)) \\
&= \tau'(\delta'((q,x), (i \bmod k)(i \div k)_k^R)) \\
&= \tau'(\delta'((q,x), (i)_k^R)),
\end{aligned}$$

concluding the proof. □

If k and n are relatively prime then the bound of Theorem 20 tends to be sharp. For instance, $\|\text{thue}\|_2^R = 2$ and $\|\text{thue}[3,0]\|_2^R = 6$ and $\|\text{thue}[5,0]\|_2^R = 10$.

If n and k are equal then Theorem 18 gives a much better bound than Theorem 20. We expect that for n and k not relatively prime, the bound of Theorem 20 is never sharp.

It is a natural question whether the same bound of Theorem 20 also holds for $\|a\|_k$ rather than $\|a\|_k^R$. The answer is negative: we have $\|\text{thue}[5,0]\|_2 = 11$ while $\|\text{thue}\|_2 = 2$. For a from Example 12 we even have $\|a[5,0]\|_2 = 12$ while $\|a\|_2 = 2$.

Apart from the operations considered in this section for many other operations on sequences the effect on the complexity measures can be investigated. Of particular interest are operations only changing the first n elements of a sequence without shifting positions. These can be obtained by combining the basic operations tail and adding initial elements, and hence yield a polynomial upper bound too. These polynomial bounds can be improved strongly as is shown in [13]. In particular, if b is obtained from a by only changing the first n elements, then $\|b\|_k \le \|a\|_k + n$ and $\|b\|_k^R \le \|a\|_k^R + \frac{kn}{k-1}$.

8. Periodic sequences

In this section we analyze both types of complexity for periodic sequences as defined in Definition 5. Recall that a sequence a is called *n-periodic* if $a_{i+n} = a_i$ for every natural number i. The set of all n-periodic sequences is denoted P_n. Note that the *period* of a (by definition the least positive integer p for which a is p-periodic) will be a divisor of n, which may be, but is not necessarily, the same as n. For an n-periodic a we write $a = (a_0 a_1 \cdots a_{n-1})^\omega$. First we give general upper bounds for both complexity measures.

Theorem 21. *Let a be an n-periodic sequence. Then* $\|a\|_k \le n$ *and* $\|a\|_k^R \le n(n-1)$.

Proof. We obtain $a_i = a_{i \bmod n}$ for all $i \in \mathbb{N}$. Define $(Q, \Sigma_k, \delta, q_0, \Gamma, \tau)$ by $Q = \{0, 1, \ldots, n-1\}$, $q_0 = 0$, $\delta(q, x) = (kq + x) \bmod n$, $\tau(q) = a_q$, for all $q \in Q, x \in \Sigma_k$. Then by induction on the length of $(i)_k$ one proves that $\delta(q_0, (i)_k) = (i \mod n)$ for every $i \in \mathbb{N}$. Hence $\tau(\delta(q_0, (i)_k)) = \tau(i \bmod n) = a_{i \bmod n} = a_i$ for all $i \in \mathbb{N}$, proving that $\|a\|_k \le n$.

For the other claim we prove that $|K_k(a)| \le n(n-1)$, then the result follows from Theorem 14. The states of $K_k(a)$ are sequences b for which there are numbers q, j such that $b_i = a_{ik^q+j} = a_{(ik^q+j) \bmod n}$ for all $i \in \mathbb{N}$. We have to show that there are at most $n(n-1)$ such sequences b. This follows from the fact that this only depends on the n values for $(j \mod n)$ and the at most $n-1$ values for $(k^q \mod n)$. The latter follows since if k, n are relatively prime, then the values of $(k^q \mod n)$ are among the $n-1$ values $1, \ldots, n-1$, and otherwise there is some $p > 1$ dividing both n and k, and the values are among the n/p multiples of p modulo n. □

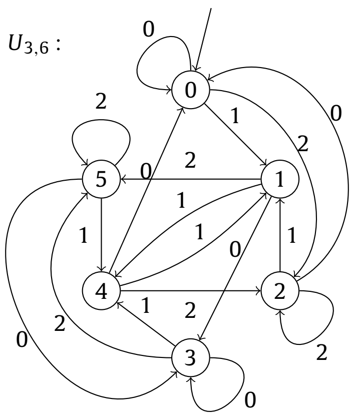
In order to improve this result and get exact values for $\|a\|_k$ and $\|a\|_k^R$ if k, n are relatively prime, we will use some properties of the integers modulo n. By $\mathbb{Z}/n\mathbb{Z}$ we denote the set of n residue classes $x \bmod n$, which becomes a group under addition modulo n. The multiplicative group modulo n is denoted by $(\mathbb{Z}/n\mathbb{Z})^*$; its elements are the residue classes $x \bmod n$ for x coprime to n, and they form a group of $\phi(n)$ elements under multiplication, where ϕ is the Euler indicator. Any residue class x in $(\mathbb{Z}/n\mathbb{Z})^*$ has a finite order $\mathrm{ord}(x, n)$ in this group, which will be a divisor of the group order $\phi(n)$, and is by definition the smallest positive integer d for which $x^d \equiv 1 \bmod n$. It is also useful to recall that ϕ is a multiplicative function on the natural numbers, with $\phi(p^k) = (p-1)p^{k-1}$ for any prime p and exponent $k \ge 1$, and in particular $\phi(n) < n$ for every $n > 1$, while $\phi(n) = n - 1$ holds for prime n only.

Since finding equivalent states, and therefore minimal DFAOs, as well as finding kernels, requires finite calculations for periodic sequences, these are well-suited for complexity calculations with a computer algebra system. The computations for which a small part is shown in this section were all done with our implementation in Magma [3].

8.1. Complexity $\|a\|_k$

The complexity of a k-automatic sequence a of period n depends on both k and n. In our first step towards determining the minimal size of a DFAO that produces $a = (a_0 a_1 \cdots a_{n-1})^\omega$ we construct a universal automaton $U_{k,n}$ of n states. The states of $U_{k,n}$ will correspond to the n different residue classes modulo n, and the transition maps are given by $\delta(x, j) = k \cdot x + j \bmod n$, for any $x \in \mathbb{Z}/n\mathbb{Z}$ and $0 \le j < k$.

Example 22. To illustrate the concept of the universal n-automaton, we exhibit $U_{3,6}$, with which we will construct some examples further on in this chapter. In the picture the states have been labeled by the residue classes, and we have not specified the output yet.



We call U the universal automaton, because any n-periodic sequence can be produced by this automaton by simply adapting the output function τ.

A minimal automaton for given a, and hence the complexity $\|a\|_k$, may now be found by merging equivalent states of $U_{k,n}$ (identifying the corresponding nodes in the directed graph) until no further equivalences exist.

As we already saw in Lemma 4, no equivalences can exist if $\gcd(k,n)=1$.

Theorem 23. *$\|a\|_k = n$ for k-automatic sequences of period n coprime to k.*

Proof. Consider the universal automaton $U_{k,n}$; let the initial state be $0 \bmod n$ and let the output function be given by $\tau(x) = a_{x \bmod n}$. It will be clear that $U_{k,n}$ will produce a: since $(k\cdot z + j)_k$ will be the concatenation of $(z)_k$ and j for $0 \le j < k$, reading the k-ary expansion of any index x will lead to the state for $x \bmod n$ and hence to output $a_{x \bmod n}$ as required. Since $U_{k,n}$ has n states, $\|a\|_k \le n$.

By Corollary 6 we know $\|a\|_k \ge n$. □

Corollary 24. *$\|tail(a)\|_k = \|a\|_k$ for any k-automatic sequence of period n coprime to k.* □

Our next goal was to relax the condition in Theorem 23 that $\gcd(k,n)=1$. Suppose that integers r and s exist such that $n = k^r \cdot s$ with s coprime to k. Note that this is equivalent to $\gcd(k,n)$ being either 1 or k. And also that this always holds if k is prime (hence in particular for 2-automatic sequences).

Although numerous computations confirm the following two results, we were not able to find a general proof, due to the intricacies of keeping track of all equivalences that arise. Therefore we present them as conjectures, generalizing Theorem 23 and Corollary 24.

Conjecture 25. *For any k-automatic sequence a of period n with $\gcd(k,n) \in \{1,k\}$:*

$$r + s \le \|a\|_k \le n,$$

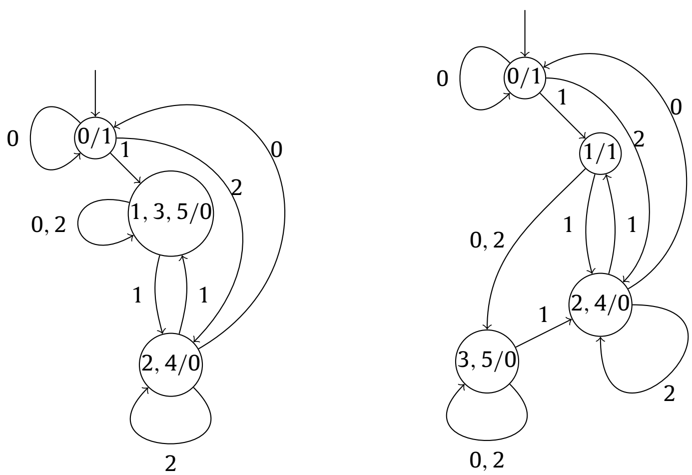
where $n = k^r \cdot s$, with s coprime to k. Moreover, for every t with $r+s \le t \le n$ there exist k-automatic sequences a of period n with $\|a\|_k = t$.

Conjecture 26. *$\|tail(a)\|_k = \|a\|_k$ for any k-automatic sequence of period n for which $\gcd(k,n) \in \{1,k\}$.*

For the remaining case, where $1 < \gcd(k,n) < k$ it is not clear exactly what the lower bound for $\|a\|_k$ should be.

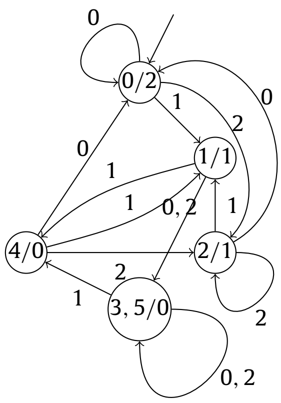
Remark 27. For any t with $r+s \le t \le n$ it is usually not difficult to exhibit a k-automatic n-periodic sequence a of complexity equal to t. For the lower bound in Conjecture 25 it should be possible to prove that $\|(10^{n-1})^\omega\|_k = r+s$, with hypotheses and notation as in the theorem, and that this period is of least complexity. The upper bound is trivially attained by the period $01\cdots n-1$ over an alphabet of n letters. Usually a considerably smaller alphabet will suffice.

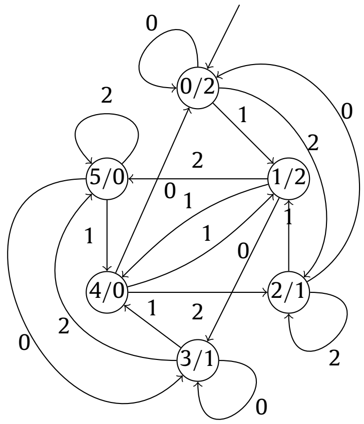
Example 28. Here we use the 6-uniform automaton from Example 22 to show that every complexity t with $3 \le t \le 6$ is assumed for 3-automatic sequences of period length 6. So in this case $k=3$, $n=6$, $r=1$ and $s=2$.



For period 100000 the complexity equals 3, as observed in Remark 27; the residues that are 1 mod 2 correspond to equivalent states that will be merged for the minimal automaton, while the class 0 mod 2 is split into the class of 0 mod 6 and the class containing both 2 mod 6 and 4 mod 6, which correspond to equivalent states. The resulting minimal automaton is shown on the left.

For period 110000 we find complexity 4, and the inequivalent states correspond to the residue classes 0 mod 6, 1 mod 6 and the pairs of equivalent states corresponding to 3, 5 mod 6 and 2, 4 mod 6, as shown on the right.





It turns out that in this case sequences of complexity 5 and 6 cannot be created over an alphabet of 2 letters. But with a 3-letter alphabet, the sequences with periods 211000 (on the left) and 221100 (with its minimal automaton shown on the right) give complexity 5, respectively 6.

8.2. Reversed complexity $\|a\|_k^R$

We now move to the reversed complexity, that is, according to Theorem 14, to determining the size of the kernel, for purely periodic sequences.

As before, denote by p_i, for $0 \le i < k$ the operation on a k-automatic sequence a of taking the subsequence $p_i(a) = (a_i, a_{i+k}, a_{i+2k} \cdots)$. The main proofs of this section use properties of the action of these operations on the set P_n of n-periodic sequences. The kernel $K_k(a)$ of any k-automatic sequence a is obtained by repeatedly applying the various p_i.

The result of applying $p_j \circ p_i$ to a will be the subsequence $(a_{i+jk+mk^2})_{m=0}^{\infty}$. Thus, the p_i under composition form a non-commutative *semigroup* $S = \langle p_i : 0 \le i < k\rangle$, with the empty product as identity element 1. Moreover, if n is coprime to k, then for each i we have $p_i^d = 1$ for $d = \mathrm{ord}(k, n)$, and d is the smallest positive integer with that property. Hence, for n coprime to k again, $p_i^{d-1} = p_i^{-1}$ and so $G = \langle p_i : 0 \le i < k\rangle$ is a finite *group*.

We derive a few more properties of the group G, in the case that k and n are coprime. First of all, one immediately checks that

$$p_1(a) = (a_1 a_{k+1} a_{2k+1} \cdots) = p_0(\mathsf{tail}(a_0 a_1 a_2 \cdots)),$$

and more generally $p_i = p_{i-1} \circ \mathsf{tail}$ for $i = 1, 2, \ldots, k-1$. But this implies on the one hand that $\mathsf{tail} = p_0^{-1} \circ p_1 \in G$, while on the other hand $p_i = p_0 \circ \mathsf{tail}^i$, and thus $p_i \in \langle p_0, \mathsf{tail}\rangle$, the subgroup of G generated by tail and p_0.

Lemma 29. *If k and n are coprime then $G = \langle p_i : 0 \le i < k\rangle$ is a group acting on the set of n-periodic sequences. Moreover, tail $\in G$ and G is generated by p_0 and tail, so $G = \langle p_0, \mathsf{tail}\rangle$, which satisfy $\mathsf{tail} \circ p_0 = p_0 \circ \mathsf{tail}^k$; finally the order of G equals $\mathrm{ord}(k, n) \cdot n$.*

Proof. Using the relations given before the statement of the theorem, every generator p_i of G can be written as $p_0 \circ \mathsf{tail}^i$, and hence every element of G as product of such elements. But the relation $\mathsf{tail} \circ p_0 = p_0 \circ \mathsf{tail}^k$ is clear from the action on any n-periodic sequence a:

$$\mathsf{tail} \circ p_0(a) = (a_k a_{2k} a_{3k} \cdots) = p_0(a_k a_{k+1} a_{k+2} \cdots) = p_0 \circ \mathsf{tail}^k(a).$$

Using this relation, every element of G can be rewritten in the form $p_0^a \circ \mathsf{tail}^b$, for certain positive integers a, b. As tail clearly has order n and p_0 has order $d = \mathrm{ord}(k, n)$ as we saw before, the proof is complete by noting that the elements $p_0^a \circ \mathsf{tail}^b$ for $0 \le a < \mathrm{ord}(k, n)$ and $0 \le b < n$ are all in G and all act differently on a sequence of period n, since tail^b acts as a cyclic permutation of the period. □

We are now in a position to draw conclusions for the reversed complexity of periodic sequences if $\gcd(k,n)=1$. We use that for periodic sequences the reversed complexity is equal to the size of the kernel, by Theorem 14.

Theorem 30. *If a is a periodic k-automatic sequence of period n, with n coprime to k, then $n \le \|a\|_k^R \le \mathrm{ord}(k,n)\cdot n$, where the upper bound is a divisor of $\phi(n)\cdot n$; in particular*

$$n \le \|a\|_k^R \le (n-1)\cdot n.$$

Proof. To compute $\|a\|_k^R$ we compute the size of the kernel $K_k(a)$. Now note that $K_k(a)$ consists of the sequences obtained by applying all possible compositions of the elements p_i, for $0 \le i < k$, to a. By Lemma 29, these elements form a group, generated by tail and p_0. Clearly, $K_k(a)$ is the size of the orbit of a^G.

If n is the period of a, the n images of $\mathrm{tail}^i(a)$ for $0 \le i < n$ will all be different, and hence the size of the kernel is at least n.

On the other hand, the size of the orbit a^G is at most $\#G$, the order of G, which is $\mathrm{ord}(k,n)\cdot n$ by Lemma 29. □

The ultimate upper bound $(n-1)\cdot n$ in Theorem 30 can only be attained for pairs k,n with n prime and k a primitive root modulo n. For every prime n such primitive root (of order $n-1$) exist, and Artin's conjecture states that a given k will be primitive root for infinitely many primes n. This conjecture has only been proved under the assumption of a generalized Riemann hypothesis.

The following theorem implies that for every $n > 8$ and every $k > 1$ the bounds n and $\mathrm{ord}(k,n)\cdot n$ in Theorem 30 will be attained for certain sequences.

Theorem 31. *Suppose that $k,n \ge 2$ and $\gcd(k,n)=1$.*

If b is the n-periodic k-automatic sequence $b=(10^{n-1})^\omega$, then $\|b\|_k^R = n$.

If c is the n-periodic k-automatic sequence $c=(10110^{n-4})^\omega$ and $n \ge 9$ then $\|c\|_k^R = \mathrm{ord}(k,n)\cdot n$.

Proof. It is clear that $\mathrm{tail}^i(b)=(0^i10^{n-i-1})^\omega$, and these give n different elements in $K_k(b)$. But with $\gcd(k,n)=1$ it will also be clear that $p_i(b)$ is n-periodic with period of 'weight' 1: there will be exactly 1 non-zero in the period. Therefore $p_i(b)$ will be among the n shifts obtained already, so $K_k(b)$ consists of n elements, for every $n \ge 2$ coprime to k.

Now consider $c=(10110^{n-4})^\omega$ for some $n \ge 9$ coprime to k. Note that

$$p_0^i(a) = a_0 a_{k^i} a_{2k^i} a_{3k^i} \cdots$$

for every positive integer i and any a. If we apply this to c, and use that it is n-periodic, we obtain that $p_0^i(c)$ is non-zero exactly at the three positions 0, $2\cdot k^i$, $3\cdot k^i$ modulo n, and $\mathrm{tail}^j p_0^i$ at positions j, $2\cdot k^i + j$, $3\cdot k^i + j$ modulo n.

The remainder of the proof consists of a verification that these triples are distinct for $0 \le j < n$ and $0 \le i < \mathrm{ord}(k,n)$, showing that the orbit c^G consists of $\mathrm{ord}(k,n)\cdot n$ different elements.

So suppose, on the contrary, that $\mathrm{tail}^j \circ p_0^i(c)$ and $\mathrm{tail}^l \circ p_0^m(c)$ coincide, that is, the sets $\{j, 2\cdot k^i + j, 3\cdot k^i + j\}$ and $\{l, 2\cdot k^m + l, 3\cdot k^m + l\}$ contain the same residues modulo n. Then these sets of residues may be permutations of each other, so we consider the following six cases (where all residues are modulo n, unless stated otherwise):

(i) $j \equiv l$, $2\cdot k^i + j \equiv 2\cdot k^{m+1} + l$, and $3\cdot k^i + j \equiv 3\cdot k^m + l$ modulo n;
in this case $j \equiv l$, $3\cdot k^i \equiv 3\cdot k^m$ and $2\cdot k^i \equiv 2\cdot k^m$, so we $k^i \equiv k^m$ and we immediately find $i \equiv m \bmod \mathrm{ord}(k,n)$. Thus the two triples are the same.

(ii) $j \equiv l$, $2\cdot k^i + j \equiv 3\cdot k^m + l$, and $3\cdot k^i + j \equiv 2\cdot k^m + l$ modulo n;
use that k is invertible modulo n, then $2\cdot k^{i-m} \equiv 3$ and $3\cdot k^{i-m} \equiv 2$ from which (by subtraction) $k^{i-m} \equiv -1$, whence $2 \equiv -3 \bmod n$, contradicting $n \ge 9$.

(iii) $j \equiv 2\cdot k^m + l$, $2\cdot k^i + j \equiv l$, and $3\cdot k^i + j \equiv 3\cdot k^m + l$ modulo n;
the first two imply that $2(k^i + k^m) \equiv 0$, while the first and third imply that $k^m \equiv 3\cdot k^i$. Together this gives $8k^i \equiv 0 \bmod n$, conflicting with the assumptions on n and k.

(iv) $j \equiv 2\cdot k^m + l$, $2\cdot k^i + j \equiv 3\cdot k^m + l$, and $3\cdot k^i + j \equiv l$ modulo n;
here the first and the second imply that $k^m \equiv 2\cdot k^i$, while first and third combine to $3\cdot k^i + 2\cdot k^m \equiv 0$. Together these yield $7\cdot k^i \equiv 0 \bmod n$, contradicting the assumptions.

(v) $j \equiv 3\cdot k^m + l$, $2\cdot k^i + j \equiv l$, and $3\cdot k^i + j \equiv 2\cdot k^m + l$ modulo n; the first and second imply $2k^i + 3k^m \equiv 0$, and the first and third give $k^m \equiv -3k^i$. But combined this implies $-7k^i \equiv 0$, which is impossible.

(vi) $j \equiv 3\cdot k^m + l$, $2\cdot k^i + j \equiv 2\cdot k^m + l$, and $3\cdot k^i + j \equiv l$ modulo n;
from the first and the second we see that $k^m \equiv -2k^i$, and from first and third that $3k^i + 3k^m \equiv 0$. Together they imply $-3k^i \equiv 0 \bmod n$, contradicting $\gcd(k,n)=1$ and $n > 8$.

We conclude that for $n > 8$ coprime to k the positions can only coincide in the very first case, when $j \equiv l \bmod n$ and $i \equiv m \bmod \mathrm{ord}(k,n)$, and thus there are $\mathrm{ord}(k,n)\cdot n$ different images in c^G. □

Example 32. Here, by way of example, is a scheme for the case that $k=2$ and $n=9$; $\text{ord}(2,9)=6=\phi(9)$ and there are 54 different elements in $K_2(c)$:

{0, 2, 3}	{1, 3, 4}	{2, 4, 5}	{3, 5, 6}	{4, 6, 7}	{5, 7, 8}	{0, 6, 8}	{0, 1, 7}	{1, 2, 8}
{0, 4, 6}	{1, 5, 7}	{2, 6, 8}	{0, 3, 7}	{1, 4, 8}	{0, 2, 5}	{1, 3, 6}	{2, 4, 7}	{3, 5, 8}
{0, 3, 8}	{0, 1, 4}	{1, 2, 5}	{2, 3, 6}	{3, 4, 7}	{4, 5, 8}	{0, 5, 6}	{1, 6, 7}	{2, 7, 8}
{0, 6, 7}	{1, 7, 8}	{0, 2, 8}	{0, 1, 3}	{1, 2, 4}	{2, 3, 5}	{3, 4, 6}	{4, 5, 8}	{0, 5, 6}
{0, 3, 5}	{1, 4, 6}	{2, 5, 7}	{3, 6, 8}	{0, 4, 7}	{1, 5, 8}	{0, 2, 6}	{1, 3, 7}	{2, 4, 8}
{0, 1, 6}	{1, 2, 7}	{2, 3, 8}	{0, 3, 4}	{1, 4, 5}	{2, 5, 6}	{3, 6, 7}	{4, 7, 8}	{0, 5, 8}

The positions where a 1 in the period appears are listed: the top left entry thus encodes the initial sequence $c=(101100000)^\omega$ and to its right all applications of tail to it. Below it c we find $p_0(c)=(100010100)^\omega$, below that $p_0^2(c)=(100100001)^\omega$ etc. It is always the case that $p_0(a)$ for a sequence a in row i can be found in row $i+1$.

About the upper bound for $n\le 8$ the following applies.

The sequence c given does also work for $n=6$, and for any k coprime to 6.

For $n=3$ and $k\equiv 2 \bmod 3$ it is necessary to enlarge the output alphabet: $(0,1,2)$ is the simplest period for which the reversed complexity 6 is attained.

For $n=4$ and $k\equiv 3 \bmod 4$ the period $(2,0,1,0)$ is the simplest case where the bound 8 for the complexity is reached.

Similarly, for $n=5$ and $k\equiv 2,3,4 \bmod 5$ the bound $\text{ord}(k,n)\cdot n$ is only attained over an alphabet of at least 3 letters.

For $n=7$ there do exist 2-automatic sequences over an alphabet of 2 letters for which the reversed complexity equals $\text{ord}(2,7)\cdot 7=21$, such as $[1,0,0,0,1,0,0]$, but the sequence c is not one of them. On the other hand, for $k=3$ reaching $\text{ord}(3,7)\cdot 7=42$ requires at least 3 letters.

For $n=8$ sequence c does attain the bound for some odd k but not for all.

Only scattered values between the lower bound n and the upper bound $\phi(n)\cdot n$ can be attained, namely only the multiples of n dividing $\phi(n)\cdot n$.

For cases where k and n are not coprime, we do not have complete results. For $k=2$ and $n=2^r s$ with $s>7$ odd, we can recursively find an upper bound $\|a\|_k^R \le \text{ord}(2,s)\cdot n+2^r-1$ (which is slightly larger than the bound in Theorem 31, see also Example 33 below). This could probably be generalized for arbitrary k dividing n; but this upper bound will not always be attained. A sharp lower bound and sharp bounds when $1<\gcd(k,n)<k$ are also lacking.

Example 33. A general strategy to create an element of $P_{2^r s}$ with maximal kernel size, is to start with 2^r 'different' elements of P_s and to use the zip operation repeatedly to create a single element of $P_{2^r s}$. The elements of P_s have to be sufficiently different to prevent any collisions under the action of S.

Let a,b,c,d, for example, be the four 9-periodic binary sequences

$$a=(110100000)^\omega, b=(111010000)^\omega, c=(111101000)^\omega, d=(111110100)^\omega$$

from P_9; each of these have the maximum size 54 for the orbit under S, much like that in Example 32. Moreover, the weights of the periods of these sequences, as of all those in their orbits, are 3, 4, 5, 6, respectively, which implies that all four orbits are disjoint. As a consequence, the orbits of the sequences $\text{zip}(a,b)$ and $\text{zip}(c,d)$ in P_{18} contain the maximum of 108 elements, and the orbit of $z=\text{zip}(\text{zip}(a,b),\text{zip}(c,d))$ contains 218 different elements, namely the previous orbits as well as $\text{zip}(a,b)$ and $\text{zip}(c,d)$ themselves. Together with z itself this gives the maximum number of 219 elements in the kernel of z.

Finally we give an illustration of the fact that (when k divides n) there exist periodic sequences a for which $\|a\|_k^R > \|a\|_k^2$.

Example 34. For the 2-automatic, 36-periodic sequence

$$a=(110000010110100010111000010111100010)^\omega$$

we find $\|a\|_2=13$ and $\|a\|_2^R=192>13^2$.

9. Conclusions and open questions

We investigated two natural complexity measures for a k-automatic sequence a: $\|a\|_k$ closely related to the alphabet size required to present a as a morphic sequence with respect to a k-uniform morphism, and $\|a\|_k^R$ closely related to the size of the kernel of a. We saw how there can be an exponential gap between $\|a\|_k$ and $\|a\|_k^R$, but basic operations like tail, adding an element in front, or applying a binary operator elementwise, never increases $\|\cdot\|_k$ or $\|\cdot\|_k^R$ by more than a quadratic factor. For combinations of these operations and many other operations the effect on the complexity measures is still open. For instance, for arithmetic subsequences $a[n,m]$ we proved $\|a[n,m]\|_k^R \leq n \cdot \|a\|_k^R$ for $0 \leq m < n$, but for a bound for $\|a[n,m]\|_k$ we only know from examples that it should be weaker. We did not investigate these measures for $m \geq n$.

For purely periodic sequences we focus on the case where the period and k are relatively prime; many open problems arise where this does not hold, like Conjectures 25, 26. It seems that reversing the period does not affect the complexity of a periodic sequence; this is certainly not true for just any permutation of the period. Also, the exact dependence of the complexity on the size of the output alphabet is not entirely clear; what is the minimal size of the alphabet to reach the complexity bounds?

Apart from these purely periodic sequences we did not consider ultimately periodic sequences; this would also be natural to consider.

Declaration of competing interest

There is no conlfict of interest.

References

[1] J.-P. Allouche, J. Shallit, Automatic Sequences: Theory, Applications, Generalizations, Cambridge University Press, 2003.
[2] D. Angluin, Learning regular sets from queries and counterexamples, Inf. Comput. 75 (2) (1987) 87–106.
[3] W. Bosma, J.J. Cannon, C. Playoust, The Magma algebra system I: the user language, J. Symb. Comput. 24 (1997) 235–265.
[4] A. Cobham, Uniform tag sequences, Math. Syst. Theory 6 (1972) 164–192.
[5] L.M. de Moura, N. Bjørner, Z3: an efficient SMT solver, in: C.R. Ramakrishnan, J. Rehof (Eds.), Tools and Algorithms for the Construction and Analysis of Systems, 14th International Conference, TACAS, in: Lecture Notes in Computer Science, vol. 4963, Springer, 2008, pp. 337–340. Tool available at https://github.com/z3prover.
[6] S. Eilenberg, Automata, Languages and Machines, Volume A, Academic Press, 1974.
[7] J. Endrullis, C. Grabmayer, D. Hendriks, Mix-automatic sequences, in: A.-H. Dediu, C. Martin-Vide, B. Truthe (Eds.), International Conference on Language and Automata Theory and Applications, LATA 2013, in: Lecture Notes in Computer Science, vol. 7810, Springer, 2013, pp. 262–274.
[8] A. Gill, Introduction to the Theory of Finite-State Machines, McGraw-Hill, 1962.
[9] G. Jiraskova, The ranges of state complexities for complement, star and reversal of regular languages, Int. J. Found. Comput. Sci. 25 (1) (2014) 101–124.
[10] M.V. Lawson, Finite Automata, Chapman and Hall/CRC, Boca Raton, FL, 2004.
[11] L. Merai, H. Niederreiter, A. Winterhof, Expansion complexity and linear complexity of sequences over finite fields, Cryptogr. Commun. 9 (4) (2017) 501–509.
[12] J. Shallit, Decidability and enumeration for automatic sequences: a survey, in: A.A. Bulatov, A.M. Shur (Eds.), 8th International Computer Science Symposium in Russia, CSR 2013, in: Lecture Notes in Computer Science, vol. 7913, Springer, 2013, pp. 49–63.
[13] P.H.M. van Spaendonck, Automatic sequences: the effect of local changes to complexity, Master thesis, Radboud University Nijmegen, 2020.
[14] H. Zantema, Complexity of automatic sequences, in: 14th International Conference on Language and Automata Theory and Applications (LATA), in: Lecture Notes in Computer Science, vol. 12038, Springer, 2020, pp. 260–271.