Complexity of Automatic Sequences

Hans Zantema[1,2(✉)]

[1] Department of Computer Science, TU Eindhoven, P.O. Box 513,
5600 MB, Eindhoven, The Netherlands
h.zantema@tue.nl
[2] Radboud University Nijmegen, P.O. Box 9010,
6500 GL, Nijmegen, The Netherlands

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 like 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 element in front, and observe that these operations may increase the complexity by at most a quadratic factor.

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] that serves as the main reference for research in this area. More recent references on the topic include [5,9].

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 Sect. 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 notation of i. The second one is similar, but then the reverse of the k-ary notation 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. We show how they relate to other characterizations; 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 sequences of automatic sequences a, b for which $\|a\|_k^R$ is exponential in $\|a\|_k$, and $\|b\|_k$ is exponential in $\|b\|_k^R$.

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 element in front rather than removing. Also other operations are considered, like pointwise combining two sequences and taking particular subsequences. About all of these basic operations f the main observation 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.

Another interesting question is what happens for periodic sequences. In the current paper we only derive a quadratic upper bound for $\|\cdot\|_k^R$ and a linear upper bound for $\|\cdot\|_k$, so opposite to the effect of tail. Whether and when these upper bounds are reached is a much more involved question that is investigated in [2]. The research project on this topic is a joined project of Wieb Bosma and the current author; as this analysis for periodic sequences requires arguments of a completely different combinatorial flavor than the automata based arguments in this paper, we decided to present the current paper and [2] separately.

Throughout the paper we make several claims about the exact values of $\|a\|_k$ and $\|a\|_k^R$ for particular sequences a. To compute these values we wrote a program to search for a DFAO of minimal size n having the corresponding property for a_i for all $i < N$ for N being typically around 2^{10}. This was done by expressing the requirements as a satisfiability problem and then call a SAT solver. The smallest n for which the formula is satisfiable then is given. As only the requirements for $i < N$ are checked, this only yields a lower bound, but for N large enough it gives the exact value. According to [6], corollary 3.1 (page 59) two states in a DFAO of n states are equivalent are equivalent if and only if for every string of length $\leq n - 1$ they produce the same output. This can be improved to $\leq n - 2$. Applying this for the union of the found automaton and the real automaton with bounds derived in this paper, this shows that for $N = 2^{n-2}$ the exact value is obtained.

This paper is organized as follows. In Sect. 2 we give the basic definitions and a general lemma for proving lower bounds. In Sect. 3 we investigate the exponential gap between $\|\cdot\|_k$ and $\|\cdot\|_k^R$. In Sect. 4 we define the kernel of an automatic sequence and investigate its relationship with $\|\cdot\|_k^R$. In Sect. 5 we present how to define automatic sequences as morphic sequences with respect to uniform morphisms, and investigate the relationship with $\|\cdot\|_k$. In Sect. 6 we investigate the effect of basic operations like tail on $\|\cdot\|_k$ and $\|\cdot\|_k^R$. In Sect. 7 we give the upper bounds of $\|\cdot\|_k$ and $\|\cdot\|_k^R$ for periodic sequences. We conclude in Sect. 8.

2   Basic Definitions

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

The set of infinite sequences $a = a_0 a_1 a_2 a_3 \cdots$ over a finite alphabet Γ is denoted by $\Gamma^{\mathbb{N}}$.

A DFA M with output (DFAO) is defined to be 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 denoted by states and arrows 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

$$(n)_k = d_0 d_1 \cdots d_r \iff n = d_0 k^r + d_1 k^{r-1} + \cdots + d_{r-1} k + d_r \wedge d_0 > 0$$

for $n > 0$. So $(0)_2 = \epsilon$ and $(11)_2 = 1011$. Note that non-empty strings of which the leftmost symbol is 0 do not occur as $(n)_k$ for some number n.

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

$$[d_0 d_1 \cdots d_r]_k = d_0 k^r + d_1 k^{r-1} + \cdots + d_{r-1} k + d_r.$$

For any Σ and any string $u \in \Sigma^*$ the reverse u^R of u is defined by $(u_1 u_2 \cdots u_n)^R = u_n u_{n-1} \cdots u_1$.

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 $M = (Q_M, \Sigma_k, \delta_M, q_0, \Gamma, \tau_M)$ 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 defined to be 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 $i \in \mathbb{N}$.*

For any k-automatic sequence $a = a_0a_1a_2a_3\cdots$ its reversed size $\|a\|_k^R$ is defined to be 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 $\|\mathsf{seq}_k^R(M)\|_k^R \leq |Q_M|$.

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 justified by the DFAO on the right.
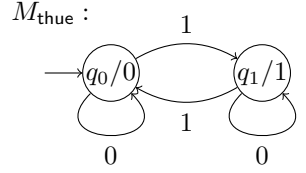
The *regular paper-folding sequence* $\mathsf{paper} = 001001100011011\cdots$ (or *dragon curve sequence* is defined by $\mathsf{paper}_i = m \mod 2$ for every $i \geq 0$ for the unique representation $i = (2m+1)2^j - 1$, see, e.g., [1] Example 5.16., or OEIS A014577. We have $\|\mathsf{paper}\|_2 = \|\mathsf{paper}\|_2^R = 4$, respectively justified by the following two DFAOs.
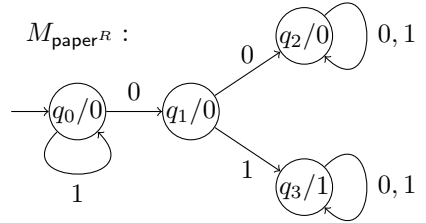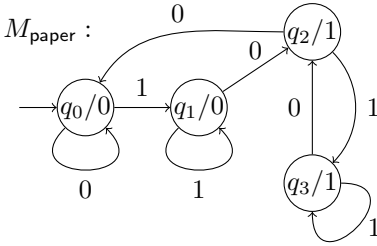
The following lemma is the basic tool for lower bounds on $\|a\|_k$ and $\|a\|_k^R$.

Lemma 1. *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 $v \in \Sigma_k^*$ satisfying $a_{[(m_i)_k v]_k} \neq a_{[(m_j)_k v]_k}$, 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 $v \in \Sigma_k^*$ satisfying $a_{[v(m_i)_k]_k} \neq a_{[v(m_j)_k]_k}$, 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 $\tau_M(\delta_M(q_i, v)) \neq \tau_M(\delta_M(q_j, v))$, so $q_i \neq q_j$. This shows $|Q| \geq n$, so $\|a\|_k \geq n$.

The proof of the second claim is similar. $\qquad\qquad\square$

3   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 [8], Sect. 3.2, page 67, exercise 3. Many similar results on state complexity are known, e.g., in [7], it is proved that all values until 2^n can be reached as sizes.

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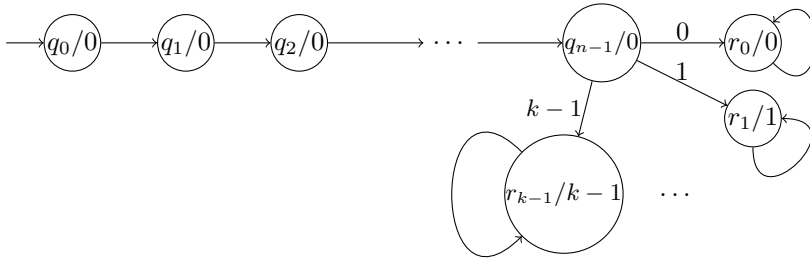
in which all unlabeled arrows are assumed to be labeled by all symbols $0, 1, \ldots, k-1$. Since this DFAO has $n+k$ states we obtain $\|a\|_k \leq n+k$.

For proving $\|a\|_k^R \geq (k-1)k^{n-1}$ we apply Lemma 1. 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 1 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.   □

4   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 [4], 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 $a = a_0 a_1 a_2 a_3 \cdots$ 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 $\tau(b_0 b_1 b_2 \cdots) = b_0$.

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 the 2-kernel exactly coincides with the DFAO M_{thue} given in Sect. 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 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 2. *For every k-automatic sequence $a = a_0 a_1 a_2 a_3 \cdots$ 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 3. *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 2 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 2 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 $M = (Q, \Sigma, \delta, q_0, \Gamma, \tau)$ such that $\tau(\delta(q_0, (i)_k^R)) = a_i$ for every $i \in \mathbb{N}$ exists. 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 $a = 01^\omega$.

Next we observe that $\|a\|_k^R$ can be strictly smaller than $|K_k(a)|$, the size of the state space of $\mathcal{K}_k(a)$. 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:
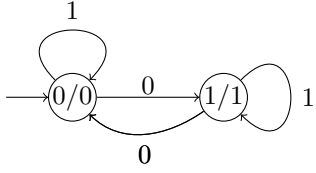
Hence $\|a\|_k^R \le 2$; we obtain $\|a\|_k^R = 2$ since the sequence contains both 0 and 1. However, $|K_k(a)| = 4$, since $\mathcal{K}_k(a)$ is the following DFAO:
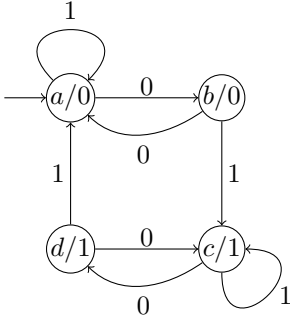
The sequences a, b, c, d are as follows:

$$a = 001001101\cdots, \quad b = 010110010\cdots,$$

$$c = 110110010\cdots, \quad d = 101001101\cdots.$$

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 then $\|a\|_k^R$.

Lemma 2. *Let a be an infinite sequence over Γ with kernel $\mathcal{K}_k(a) = (K_k(a), \Sigma_k, \delta, a, \Gamma, \tau)$. Let $(Q_M, \Sigma_k, \delta_M, q_0, \Gamma, \tau_M)$ such that $\tau_M(\delta_M(q_0, (i)_k^R)) = a_i$ for all $i \in \mathbb{N}$. 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: $\quad \delta(a, u)_i = a_{m_u} = \tau_M(\delta_M(q_0, (m_u)_k^R))$
$$\begin{aligned}
&= \tau_M(\delta_M(q_0, u(i)_k^R)) \\
&= \tau_M(\delta_M(\delta_M(q_0, u), (i)_k^R)) \\
&= \tau_M(\delta_M(\delta_M(q_0, v), (i)_k^R)) \\
&= \tau_M(\delta_M(q_0, (m_v)_k^R)) = a_{m_v} = \delta(a, v)_i.
\end{aligned}$$
$\square$

Theorem 4. *Let a be a k-automatic sequence over an alphabet Γ. Then*

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

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 2 $b \sim c$ 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 $|K_k(a)| \leq |\Gamma| * \|a\|_k^R$.

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 $\|a\|_k^R = |K_k(a)|$. □

5   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 4, $\|a\|_k$ is strongly related to the number of symbols needed to describe a as a morphic sequence with respect to a k-uniform morphism. A sequence a over an alphabet Γ is called *morphic* with respect to a morphism $h : \Delta \rightarrow \Delta^*$ and a coding $\tau : \Delta \rightarrow \Gamma$ if $a = \tau(h^\omega(x))$ for some $x \in \Delta$ satisfying $h(x) = xu$, by which $h^\omega(x) = xuh(u)h^2(u)h^3(u)\cdots$ is a fixed point of h. The morphism $h : \Delta \rightarrow \Delta^*$ is called k-uniform if the string $h(y) \in \Delta^*$ has length k for every $y \in \Delta$. It is well-known (Cobham [3], 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, thue $= h^\omega(0)$ for $h(0) = 01, h(1) = 10$, and 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 5. *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 $h : \Delta \rightarrow \Delta^*$ and a coding $\tau : \Delta \rightarrow \Gamma$. Then $\|a\|_k \leq d(a) \leq \|a\|_k + 1$.*

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 $i \geq 0$ 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 $\|a\|_k \leq d(a)$.

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 $i \geq 0$, 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 $y \neq 0$, $\tau(q_0') = \tau_M(q_0)$, $\tau(q) = \tau_M(q)$ 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 $i \geq 0$. 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 $d(a) \leq \|a\|_k + 1$. □

6 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 6. *For any k-automatic sequence a we have $\|\mathsf{tail}(a)\|_k^R \leq 2\|a\|_k^R$ and $\|\mathsf{tail}(a)\|_k \leq (\|a\|_k)^2$. 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 $\|a\|_k^R$ with $\tau(\delta(q_0, (i)_k^R)) = a_i$ for all $i \geq 0$. Let $m \leq \|a\|_k^R$ be the smallest number $m > 0$ such that $j < m$ exists with $\delta(q_0, 0^m) = \delta(q_0, 0^j)$. 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 $|Q \cup \{r_0, \ldots, r_{m-1}\}| \leq 2|Q|$ this yields $\|\mathsf{tail}(a)\|_k^R \leq 2\|a\|_k^R$.

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 $i \geq 0$. Define the DFAO $\overline{M} = (Q \times Q, \Sigma_k, \overline{\delta}, \overline{q_0}, \Gamma, \overline{\tau})$ of size $(\|a\|_k)^2$ by

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

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

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

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 $m \geq 0, v \in \Sigma_k^*, x < k-1$. 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 $\|\mathsf{tail}(a)\|_k \leq n^2$, for the last claim it suffices to prove $\|\mathsf{tail}(a)\|_k \geq n^2$. 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 $\|a\|_k \leq n$, 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 $\|\mathsf{tail}(a)\|_k \geq n^2$ by Lemma 1. 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 $i \neq j$, then $(p, q) \neq (p', q')$.

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 $b_{[(m_i)_k v]_k} = a_{[(m_i)_k v]_k + 1} = a_{[10^{p-1}10^{q+r}]_k} = 1 \neq 0 = a_{[10^{p'-1}10^{q'+r}]_k} = b_{[(m_j)_k v]_k}$.

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 1 hold, and $\|\mathsf{tail}(a)\|_k \geq n^2$. $\qquad\square$

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

For any sequence $a = a_0 a_1 a_2 a_3 \cdots$ over Γ, and $x \in \Gamma$ the sequence $x \cdot a = x a_0 a_1 a_2 a_3 \cdots$ 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 $x\cdot$ is similar to tail.

Theorem 7. *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 6, 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 $x > 0$. $\qquad\square$

For our examples thue and paper we have $\|0 \cdot \mathsf{thue}\|_2 = 4$, $\|0 \cdot \mathsf{thue}\|_2^R = 4$, $\|0 \cdot \mathsf{paper}\|_2 = 4$ and $\|0 \cdot \mathsf{paper}\|_2^R = 4$.

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 8. *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 $\|p_j(a)\|_k \leq \|a\|_k$. $\qquad\square$

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

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.

Theorem 9. *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 $\|f(a,b)\|_k \le \|a\|_k \|b\|_k$ and $\|f(a,b)\|_k^R \le \|a\|_k^R \|b\|_k^R$.*

Proof. Let $(Q_1, \Sigma_k, \delta_1, q_{10}, \Gamma_1, \tau_1)$ be a DFAO of size $\|a\|_k$ with $\tau_1(\delta(q_{10}, (i)_k)) = a_i$ for all $i \ge 0$. Let $(Q_2, \Sigma_k, \delta_2, q_{20}, \Gamma_2, \tau_2)$ be a DFAO of size $\|b\|_k$ with $\tau_2(\delta(q_{20}, (i)_k)) = b_i$ for all $i \ge 0$. 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 $\|a\|_k \|b\|_k$ for $f(a,b)$. The proof for the reversed version is similar. □

Combining our examples thue and paper we have $\|\mathsf{thue} \wedge \mathsf{paper}\|_2 = 8$ and $\|\mathsf{thue} \wedge \mathsf{paper}\|_2^R = 7$.

7   Periodic Sequences

Theorem 10. *Let $a = v^{\omega}$ be a periodic sequence with $|v| = n$. Then $\|a\|_k \le n$ and $\|a\|_k^R \le n(n-1)$.*

Proof. Writing $v = v_0 v_1 \cdots v_{n-1}$ we obtain $a_i = v_{i \bmod n}$ for all $i \in \mathbb{N}$. Define $(Q, \Sigma_k, \delta, q_0, \Gamma, \tau)$ by $Q = \{0, 1, \ldots, n-1\}$, $q_0 = 0$, $\delta(q, x) = (kq + x) \bmod n$, $\tau(q) = v_q$, for all $q \in Q, x \in \Sigma_k$. Then by induction on the length of $(i)_k$ one proves that $\delta(q_0, (i)_k) = (i \bmod n)$ for every $i \in \mathbb{N}$. Hence $\tau(\delta(q_0, (i)_k)) = \tau(i \bmod n) = v_{i \bmod n} = a_i$ 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 4. The states of $K_k(a)$ are sequences b for which there are numbers q, j such that $b_i = a_{ik^q + j} = v_{(ik^q + j) \bmod n}$ 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 i depends on the n values for $(j \bmod n)$ and the at most $n-1$ values for $(k^q \bmod n)$. The latter follows since if k, n are relatively prime, then the values of $(k^q \bmod n)$ 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. □

A natural question is for which cases the bounds of Theorem 10 can be reached, in particular the quadratic bound for $\|a\|_k^R$. This question is beyond the scope of this paper, but has been addressed in [2]. A main result of [2] is that if $n > 5$ is prime and 2 is a primitive root modulo n (on which Artin's conjecture states that this holds for infinitely many primes), then $\|v^{\omega}\|_k^R = n(n-1)$ for $v = 10110^{n-4}$.

8   Conclusions

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. Many other operations, like changing the tenth element of a sequence, can be obtained by combining such basic operations, and hence yield a polynomial upper bound too. Probably these polynomial bounds can be improved strongly. Other open questions include a further investigation of when these upper bounds can be reached. Conversely, our SAT based tool provides values that are likely to be exact, but formally are only lower bounds. It would make sense to further investigate how to be sure to have the exact value, either depending on particular ways to define automatic sequences, or by giving general criteria for exactness depending on known upper bounds.

On periodic sequences this paper only contains some very basic observations; more involved observations are given in [2].

We want to thank Wieb Bosma for fruitful collaboration on this topic and careful proof reading. We want to thank Jeffrey Shallit for giving pointers to state complexity.

References

1. Allouche, J.P., Shallit, J.: Automatic Sequences: Theory, Applications, Generalizations. Cambridge University Press, Cambridge (2003)
2. Bosma, W.: Complexity of periodic sequences (2019). https://www.math.ru.nl/~bosma/pubs/periodic.pdf
3. Cobham, A.: Uniform tag sequences. Math. Systems Theory **6**, 164–192 (1972)
4. Eilenberg, S.: Automata, Languages and Machines, vol. A. Academic Press, New York (1974)
5. Endrullis, J., Grabmayer, C., Hendriks, D.: Mix-automatic sequences. In: Dediu, A.-H., Martín-Vide, C., Truthe, B. (eds.) LATA 2013. LNCS, vol. 7810, pp. 262–274. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-37064-9_24
6. Gill, A.: Introduction to the Theory of Finite-State Machines. McGraw-Hill, New York (1962)
7. Jiraskova, G.: The ranges of state complexities for complement, star and reversal of regular languages. Int. J. Found. Comput. Sci. **25**(1), 101–124 (2014)
8. Lawson, M.V.: Finite Automata. Chapman and Hall/CRC, Boca Raton (2004)
9. Shallit, J.: Decidability and enumeration for automatic sequences: a survey. In: Bulatov, A.A., Shur, A.M. (eds.) CSR 2013. LNCS, vol. 7913, pp. 49–63. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-38536-0_5