HAL open science

Operations on Weakly Recognizing Morphisms

Lukas Fleischer, Manfred Kufleitner

To cite this version:

Lukas Fleischer, Manfred Kufleitner. Operations on Weakly Recognizing Morphisms. 18th International Workshop on Descriptional Complexity of Formal Systems (DCFS), Jul 2016, Bucharest, Romania. pp.126-137, 10.1007/978-3-319-41114-9_10 . hal-01633955

HAL Id: hal-01633955

https://inria.hal.science/hal-01633955

Submitted on 13 Nov 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Operations on Weakly Recognizing Morphisms

Lukas Fleischer and Manfred Kufleitner

FMI, University of Stuttgart*
Universitätsstraße 38, 70569 Stuttgart, Germany
{fleischer,kufleitner}@fmi.uni-stuttgart.de

Abstract. Weakly recognizing morphisms from free semigroups onto finite semigroups are a classical way for defining the class of ω-regular languages, i.e., a set of infinite words is weakly recognizable by such a morphism if and only if it is accepted by some Büchi automaton. We consider the descriptional complexity of various constructions for weakly recognizing morphisms. This includes the conversion from and to Büchi automata, the conversion into strongly recognizing morphisms, and complementation. For some problems, we are able to give more precise bounds in the case of binary alphabets or simple semigroups.

1 Introduction

Büchi automata define the class of ω-regular languages. They were introduced by Büchi for deciding the monadic second-order theory of $(\mathbb{N}, <)$ [2]. Since then, ω-regular languages have become an important tool in formal verification, and many other automata models for this language class have been considered; see e.g. [10, 13]. Each automaton model has its merits and its disadvantages. Recently, the authors have shown that recognizing morphisms have many nice algorithmic properties [5]. Such morphisms come in two different flavors. Strongly recognizing morphisms admit efficient minimization and complementation, whereas weakly recognizing morphisms can be exponentially more succinct (but there is no minimal weak recognizer and there is no efficient complementation). The situation is similar to the behavior of deterministic and nondeterministic finite automata. The major difference to both nondeterministic finite automata and Büchi automata is that there is an efficient inclusion test for weakly recognizing morphisms [5]. Every strongly recognizing morphism is also weakly recognizing, but the converse is false.

In this paper, we consider the descriptional complexity of various operations on weakly recognizing morphisms and conversions involving nondeterministic Büchi automata (BA) and strongly recognizing morphisms. In each case, we give asymptotically tight bounds. For the conversion of a BA into a weakly recognizing morphism, we give a lower bound which matches the naive upper bound. Our results are summarized in Table 1.

There are some similarities between recognizing morphisms over finite and over infinite words. Strong recognition is the natural counterpart to recognition

* This work was supported by the DFG grants DI 435/5-2 and KU 2716/1-1.

Operation	Lower bound	Upper bound
BA to weak recognition	2^{n^2} [**new**]	2^{n^2} [9]
BA to weak recognition, binary alphabet	$2^{(n-1)^2/4}$ [**new**]	2^{n^2} [9]
Weak recognition to BA	$(n-3)(n+1)/32$ [**new**]	$n(n+1)$ [9]
Weak recognition to strong recognition	$n2^{n-1}$ [**new**]	2^{n^2} [10]
Complementation of weak recognition	$n2^{n-1}$ [**new**]	2^{n^2} [10]
Complementation for simple semigroups	$n2^{n-1}$ [**new**]	$n2^n$ [**new**]

Table 1. Bounds for the descriptional complexity of various operations.

for finite words. Nevertheless, in order to prove lower bounds for the conversion of Büchi automata to weakly recognizing morphisms, we first show that bounds for converting nondeterministic finite automata to recognizing morphisms over finite words (with some limitations) also hold for the conversion of Büchi automata to weakly recognizing morphisms. We then use techniques of Sakoda and Sipser [12] and of Yan [14] to obtain tight bounds for the conversion of nondeterministic finite automata to recognizing morphisms. This step is similar to the work of Holzer and König [6]. To the best of our knowledge, our lower bound over finite words for the conversion of an NFA into a recognizing morphism is also a new result.

2 Preliminaries

This section gives a brief overview of some basic definitions from the fields of formal languages, finite automata and semigroup theory. We refer to [10, 11] for more detailed introductions.

Words. Let A be a finite *alphabet*. The elements of A are called *letters*. A *finite word* is a sequence $a_1a_2\cdots a_n$ of letters of A and an *infinite word* is an infinite sequence $a_1a_2\cdots$. The empty word is denoted by ε. Given an infinite word $\alpha = a_1a_2\cdots$, we let $\inf(\alpha) \subseteq A$ denote the set of letters in α which occur infinitely often.

Let K be a set of finite words and let L be a set of infinite words. We set $KL = \{u\alpha \mid u \in K, \alpha \in L\}$, $K^n = \{u_1u_2\cdots u_n \mid u_i \in K\}$, $K^+ = \bigcup_{n\geqslant 1} K^n$ and $K^* = K^+ \cup \{\varepsilon\}$. Moreover, if $\varepsilon \notin K$ we define the *infinite iteration* $K^\omega = \{u_1u_2\cdots \mid u_i \in K\}$. A natural extension to $K \subseteq A^*$ is $K^\omega = (K \setminus \{\varepsilon\})^\omega \cup \{\varepsilon\}$.

Automata. A *finite automaton* is a 5-tuple $\mathcal{A} = (Q, A, \delta, I, F)$ where Q is a finite set of *states* and A is a finite alphabet. The *transition relation* δ is a subset of $Q \times A \times Q$ and its elements are called *transitions*. The sets I and F are subsets of Q and are called *initial states* and *final states*, respectively.

A *finite run* of a word $a_1a_2\cdots a_n$ on $\mathcal{A}$ is a sequence $q_0a_1q_1a_1\cdots q_{n-1}a_nq_n$ such that $q_0 \in I$ and $(q_i, a_{i+1}, q_{i+1}) \in \delta$ for all $i \in \{0, \ldots, n-1\}$. The run is said

to *start* in q_0 and *end* in q_n. The word $a_1 a_2 \cdots a_n$ is the *label* of the run. A finite run is called *accepting* if it ends in a final state. A finite word u is said to be *accepted by* $\mathcal{A}$ if there exists an accepting finite run of u on $\mathcal{A}$ and the language *accepted by* $\mathcal{A}$ is the set of all finite words over A^* accepted by $\mathcal{A}$. It is denoted by $L_{\mathrm{NFA}}(\mathcal{A})$.

Analogously, an *infinite run* of a word $a_1 a_2 \cdots$ on $\mathcal{A}$ is an infinite sequence $q_0 a_1 q_1 a_1 \cdots$ such that $q_0 \in I$ and $(q_i, a_{i+1}, q_{i+1}) \in \delta$ for all $i \geqslant 0$. It is called *accepting* if $\inf(q_0 q_1 q_2 \cdots) \cap F \neq \emptyset$. An infinite word α is said to be *Büchi-accepted by* $\mathcal{A}$ if there exists an accepting infinite run of α on $\mathcal{A}$. The language *Büchi-accepted by* $\mathcal{A}$ is the set of all infinite words Büchi-accepted by $\mathcal{A}$ and it is denoted by $L_{\mathrm{BA}}(\mathcal{A})$.

We use the term *run* for both finite and infinite runs if the reference is clear from the context. A language $L \subseteq A^*$ (resp. $L \subseteq A^\omega$) is *regular* (resp. *ω-regular*) if it is accepted (resp. Büchi-accepted) by some finite automaton.

Finite semigroups. A *semigroup morphism* is a mapping $h\colon S \to T$ between two (not necessarily finite) semigroups S and T such that $h(s)h(t) = h(st)$ for all $s, t \in S$. Since we do not consider morphisms of other objects, we use the term *morphism* synonymously. A *subsemigroup* of a semigroup S is a subset that is closed under multiplication. We say that a semigroup T *divides* a semigroup S if there exists a surjective morphism from a subsemigroup of S onto T.

Green's relations are an important tool in the study of semigroups. For the remainder of this subsection, let S be a finite semigroup. We let S^1 denote the monoid that is obtained by adding a new neutral element 1 to S. For $s, t \in S$ let

$$
\begin{aligned}
&s \mathrel{\mathcal{R}} t \text{ if there exist } q, q' \in S^1 \text{ such that } sq = t \text{ and } tq' = s,\\
&s \mathrel{\mathcal{L}} t \text{ if there exist } p, p' \in S^1 \text{ such that } ps = t \text{ and } p't = s,\\
&s \mathrel{\mathcal{J}} t \text{ if there exist } p, q, p', q' \in S^1 \text{ such that } psq = t \text{ and } p'tq' = s,\\
&s \mathrel{\mathcal{H}} t \text{ if } s \mathrel{\mathcal{R}} t \text{ and } s \mathrel{\mathcal{L}} t.
\end{aligned}
$$

These relations are equivalence relations. The equivalence classes of $\mathcal{R}$ (resp. $\mathcal{L}$, $\mathcal{J}$, $\mathcal{H}$) are called *$\mathcal{R}$-classes* (resp. *$\mathcal{L}$-classes*, *$\mathcal{J}$-classes*, *$\mathcal{H}$-classes*). For $s \in S$, we denote the $\mathcal{R}$-class (resp. $\mathcal{L}$-class) of s by R_s (resp. L_s) and we let $S/\mathcal{R} = \{R_s \mid s \in S\}$ as well as $S/\mathcal{L} = \{L_s \mid s \in S\}$.

A semigroup is called *$\mathcal{J}$-trivial* if each of its $\mathcal{J}$-classes contains exactly one element. A semigroup is called *simple* if it consists of a single $\mathcal{J}$-class. In a finite simple semigroup, the relations $s \mathrel{\mathcal{R}} st \mathrel{\mathcal{L}} t$ hold for all $s, t \in S$. Moreover, each $\mathcal{H}$-class forms a group and all such groups are isomorphic [11]. We will also utilize the following lemma:

Lemma 1. *Let S be a finite simple semigroup and let $x, y, z \in S$ such that $y \mathrel{\mathcal{R}} z$. Then $xy = xz$ implies $y = z$.*

Proof. Suppose that $xy = xz$. Since S is simple, we have $y \mathrel{\mathcal{L}} xy$ and thus, there exists an element $p \in S^1$ such that $pxy = y$. Since $y \mathrel{\mathcal{R}} z$, there exists an element $q \in S^1$ with $yq = z$. It follows that $y = pxy = pxz = pxyq = yq = z$. $\square$

Recognition by morphisms. Let $h\colon A^+ \to S$ be a morphism to a finite semigroup S. A pair (s, e) of elements of S is a *linked pair* if $se = s$ and $e^2 = e$. For $s \in S$, we set $[s]_h = h^{-1}(s)$ and if h is understood from the context, we may skip the reference to the morphism in the subscript. A language $L \subseteq A^+$ is *recognized* by a morphism $h : A^+ \to S$ if L is a union of sets $[s_i]$ with $s_i \in S$. A language $L \subseteq A^\omega$ is *weakly recognized* by a morphism $h : A^+ \to S$ if it is a union of sets $[s_i][e_i]^\omega$ where (s_i, e_i) are linked pairs of S. A language $L \subseteq A^\omega$ is *strongly recognized* by a morphism $h : A^+ \to S$ if $[s][t]^\omega \cap L \neq \emptyset$ implies $[s][t]^\omega \subseteq L$ for all $s, t \in S$. It is easy to see that strong recognition implies weak recognition, see e.g. [10, Theorem 2.2]. Moreover, if a morphism strongly recognizes L, it also strongly recognizes its complement $A^\omega \setminus L$. By extension, we also say that a semigroup S recognizes (resp. weakly recognizes, strongly recognizes) a language L if there exists a morphism $h\colon A^+ \to L$ that recognizes (resp. weakly recognizes, strongly recognizes) L.

For a language $L \subseteq A^+ \cup A^\omega$, we have $u \equiv_L v$ if and only if

$$(xuy)z^\omega \in L \Leftrightarrow (xvy)z^\omega \in L \text{ and}$$
$$z(xuy)^\omega \in L \Leftrightarrow z(xvy)^\omega \in L$$

for all finite words $x, y, z \in A^*$. Keep in mind that $\varepsilon^\omega = \varepsilon$. The relation $\equiv_L$ was introduced by Arnold [1]; it is called the *syntactic congruence* of L. The congruence classes of $\equiv_L$ form the so-called *syntactic semigroup* $A^+/{\equiv_L}$ and the *syntactic morphism* $h_L\colon A^+ \to A^+/{\equiv_L}$ is the natural quotient map. If $L \subseteq A^*$ (resp. $L \subseteq A^\omega$) is regular (resp. ω-regular), the syntactic semigroup of L is finite and h_L recognizes (resp. strongly recognizes) the language L; see [1, 10].

3 Lower Bound Techniques

3.1 Proving Lower Bounds for Weakly Recognizing Morphisms

We first consider the general problem of proving lower bounds for the size of weakly recognizing semigroups for a given language L. In the case of recognizing morphisms over finite words and in the case of strongly recognizing morphisms, this is easy since one only needs to compute the syntactic semigroup, which immediately yields a tight lower bound. On the contrary, weakly recognizing morphisms do not admit minimal objects. However, it turns out that one can still use a relaxed version of Arnold's syntactic congruence.

We first prove a combinatorial lemma and then give the main result of this section.

Lemma 2. *Let $u, v \in A^+$ and let (s, e) be a linked pair. Then uv^ω is contained in $[s][e]^\omega$ if and only if there exists a factorization $v = v_1 v_2$ and powers $k, \ell \geqslant 0$ such that ℓ is odd, $h(uv^k v_1) = s$ and $h(v_2 v^\ell v_1) = e$.*

Proof. Let $v = a_1 a_2 \cdots a_n$ with $n \geqslant 1$ and $a_i \in A$. If uv^ω is contained in $[s][e]^\omega$, there exists a factorization $uv^\omega = u' v'_1 v'_2 \cdots$ such that $h(u') = s$ and $h(v'_i) = e$

for all $i \geqslant 1$. Since u and v are finite words, there exist indices $j > i \geqslant 1$, powers $k, \ell \geqslant 1$ and a position $m \in \{1, \ldots, n\}$ such that $u'v'_1v'_2 \cdots v'_{i-1} = uv^k a_1 a_2 \cdots a_m$ and $v'_i v'_{i+1} \cdots v'_j = a_{m+1}a_{m+2} \cdots a_n v^\ell a_1 a_2 \cdots a_m$. We set $v_1 = a_1 a_2 \cdots a_m$ and $v_2 = a_{m+1}a_{m+2} \cdots a_n$. Then $v_1 v_2 = v$,

$$\begin{aligned} h(uv^k v_1) &= h(uv^k a_1 a_2 \cdots a_m) &&= h(u'v'_1v'_2 \cdots v'_{i-1}) = se^{i-1} &&= s, \\ h(v_2 v^\ell v_1) &= h(a_{m+1}a_{m+2} \cdots a_n v^\ell a_1 a_2 \cdots a_m) &&= h(v'_i v'_{i+1} \cdots v'_j) &&= e^{j-i+1} = e. \end{aligned}$$

If ℓ is even, we can replace ℓ by $2\ell + 1$ since $h(v_2 v^{2\ell+1} v_1) = h(v_2 v^\ell v_1 v_2 v^\ell v_1) = e^2 = e$. The converse implication is trivial. □

Theorem 3. *Let $L \subseteq A^\omega$ be a language weakly recognized by some morphism $h\colon A^+ \to S$ and let $u, v, z \in A^+$ and $x, y \in A^*$ be words such that one of the following two properties holds:*

1. *$xuyz^\omega \in L$ and $xvyz^\omega \notin L$*
2. *$x(uy)^\omega \in L$ and $x(uyvy)^\omega \notin L$ and $x(vyuy)^\omega \notin L$.*

Then $h(u) \neq h(v)$.

Proof. We consider finite words $u, v \in A^+$ such that $h(u) = h(v)$ and show that in this case, neither of the properties can hold.

If the first property holds, there exists a linked pair (s, e) such that $xuyz^\omega \in [s][e]^\omega \subseteq L$. Thus, by Lemma 2, we have $h(xuyz^k z_1) = s$ and $h(z_2 z^\ell z_1) = e$ for some factorization $z = z_1 z_2$ and powers $k, \ell \geqslant 0$. Now, since $h(xvyz^k z_1) = h(xuyz^k z_1) = s$, we obtain $xvyz^\omega \in [s][e]^\omega \subseteq L$, a contradiction.

If the second property holds, there exists a linked pair (s, e) of S such that $xw^\omega \in [s][e]^\omega \subseteq L$ where $w = uy$. Thus, by Lemma 2, we have $h(xw^k w_1) = s$ and $h(w_2 w^\ell w_1) = e$ for some factorization $w = w_1 w_2$, some power $k \geqslant 0$ and some odd power $\ell \geqslant 0$. Since ℓ is odd $(\ell - 1)/2$ is an integer and we have $h(w_2 (vyuy)^{(\ell-1)/2} vyw_1) = h(w_2 (uy)^\ell w_1) = e$. Now, if k is odd as well, we obtain $h(x(vyuy)^{(k-1)/2} vyw_1) = h(x(uy)^k w_1) = s$ and therefore, $x(vyuy)^\omega \in L$. Equivalently, if k is even, we have $h(x(uyvy)^{k/2} w_1) = h(x(uy)^k w_1) = s$ and hence, $x(uyvy)^\omega \in L$. Both cases contradict Property 2 above. □

The next proposition is another simple, yet useful, tool for proving lower bounds. It allows to transfer bounds from the setting of finite words to infinite words.

Proposition 4. *Let $\mathcal{A} = (Q, A, \delta, I, F)$ and let $a \in A$ be a letter such that for all $q \in Q$ and $q_f \in F$, we have $(q, a, q_f) \in \delta$ if and only if $q = q_f$. Let $K = L_{BA}(\mathcal{A})$ and let $L = L_{NFA}(\mathcal{A})$. Then each semigroup weakly recognizing K has at least $|A^+/{\equiv_L}|$ elements.*

Proof. Let $h\colon A^+ \to S$ be a morphism weakly recognizing K and consider two words $u, v \in A^+$ such that $u \not\equiv_L v$. Then, without loss of generality, there exist $x, y \in A^*$ such that $xuy \in L$ and $xvy \notin L$. This implies $xuya^\omega \in K$ since $(q_f, a, q_f) \in \delta$ for all $q_f \in F$. Equivalently, because of $(q, a, q_f) \notin \delta$ for all $q \in Q \setminus F$ and $q_f \in F$, we have $xvya^\omega \notin K$. By Theorem 3, this yields $h(u) \neq h(v)$. □

3.2 The Full Automata Technique

The *full automata technique* is a useful tool for proving lower bounds for the conversion of automata to other objects. It was introduced by Yan [14] who attributes it to Sakoda and Sipser [12]. The technique works for both accepted and Büchi-accepted languages. However, we will prove the main result of this section only for the setting of finite words and use Proposition 4 to obtain analogous results for infinite words.

Let Q be a finite set and let I, F be subsets of Q. The *full automaton* $\mathcal{F}(Q, I, F)$ is the finite automaton (Q, B, Δ, I, F) defined by $B = 2^{Q^2}$ and by the transition relation $\Delta = \{(p, T, q) \in Q \times B \times Q \mid (p, q) \in T\}$.

Theorem 5. *Let $\mathcal{A} = (Q, A, \delta, I, F)$ be a finite automaton and let $\mathcal{F}(Q, I, F) = (Q, B, \Delta, I, F)$ be the corresponding full automaton. Then the syntactic semigroup of $L_{NFA}(\mathcal{A})$ divides the syntactic semigroup of $L_{NFA}(\mathcal{F}(Q, I, F))$.*

Proof. We first define a morphism $\pi\colon A^+ \to B^+$ by $\pi(a) = \{(p, q) \mid (p, a, q) \in \delta\}$. Let $K = L_{\mathrm{NFA}}(\mathcal{F}(Q, I, F))$ and let $L = L_{\mathrm{NFA}}(\mathcal{A})$. It suffices to show that $\pi(u) \equiv_K \pi(v)$ implies $u \equiv_L v$. Thus, consider $u, v \in A^+$ such that $\pi(u) \equiv_K \pi(v)$. In particular, for all $x, y \in A^*$, we have $\pi(xuy) \in K$ if and only if $\pi(xvy) \in K$. By the definition of π, we have $\pi(w) \in K$ if and only if $w \in L$ for all $w \in A^+$. Using the equivalence from above, this yields $xuy \in L$ if and only if $xvy \in L$ for all $x, y \in A^*$, thereby proving that $u \equiv_L v$. □

4 From Automata to Weakly Recognizing Morphisms

The standard construction for converting a finite automaton $\mathcal{A}$ to a recognizing morphism is the so-called *transition semigroup* of $\mathcal{A}$. For a given word $u \in A^+$, it encodes for each pair (p, q) of states whether there is a run of u on $\mathcal{A}$ starting in p and ending in q. Thus, for a finite automaton with n states the transition semigroup has 2^{n^2} elements. For details on the construction, we refer to [10, 11]. We show that this construction is optimal.

Theorem 6. *Let $\mathcal{A}$ be a finite automaton with n states. Then there exists a semigroup recognizing $L_{NFA}(\mathcal{A})$ (resp. weakly recognizing $L_{BA}(\mathcal{A})$) which has at most 2^{n^2} elements and this bound is tight.*

Proof. Each language that is accepted (resp. Büchi-accepted) by $\mathcal{A}$ is recognized (resp. weakly recognized) by the transition semigroup of $\mathcal{A}$ which has size 2^{n^2}.

To show that this is optimal, we consider the full automaton $\mathcal{F}(N, N, N) = (N, B, \Delta, N, N)$ where $N = \{1, \ldots, n\}$ and let $L = L_{\mathrm{NFA}}(\mathcal{F}(N, N, N))$. For two different letters $X, Y \in B$ we may assume, without loss of generality, that there exist $p, q \in N$ such that $(p, q) \in X \setminus Y$. With $P = \{(p, p)\}$ and $Q = \{(q, q)\}$, we then have $PXQ \in L$ and $PYQ \notin L$. Thus, $X \not\equiv_L Y$. This shows that $B^+/{\equiv_L}$ has at least $|B| = 2^{n^2}$ elements.

Noting that the transitions labeled by the letter $\{(q, q) \mid q \in N\}$ form self-loops at each state, the Büchi case immediately follows by Proposition 4. □

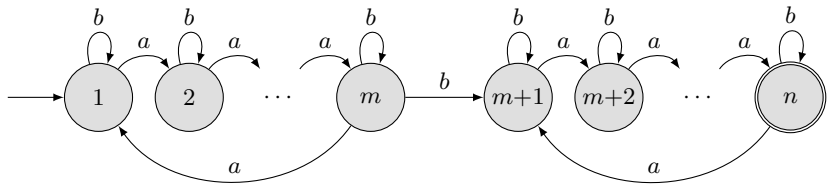
The proof of the optimality result requires a large alphabet that grows super-exponentially in the number of states of the automaton. A natural restriction is considering automata over fixed-size alphabets.

By a result of Chrobak [3], the size of the syntactic semigroup of an unary language accepted by a finite automaton of size n is in $2^{\mathcal{O}(\sqrt{n \log n})}$ (note that since unary languages are commutative, the syntactic monoid is isomorphic to the minimal deterministic automaton). Over infinite words, the unary case is uninteresting since the only language over the alphabet $A = \{a\}$ is $\{a^\omega\}$.

For binary alphabets, a lower bound can be obtained by combining the full automata technique with a result from the study of semigroups of binary relations [7, Proposition 6]. In order to keep the paper self-contained, we present a proof that is adapted to finite automata and does not require any knowledge of binary relations.

Theorem 7. *Let $A = \{a, b\}$ and let n be an odd natural number. There exists a language $L \subseteq A^+$ (resp. $L \subseteq A^\omega$) and a finite automaton with n states accepting (resp. Büchi-accepting) L, such that each semigroup recognizing (resp. weakly recognizing) L has at least $2^{(n-1)^2/4}$ elements.*

Proof. We first analyze the case of finite words. Let $m = (n-1)/2$ and let $M = \{1, \ldots, m\}$. We consider the automaton $\mathcal{A}$ depicted below and let $L = L_{\mathrm{NFA}}(\mathcal{A})$.



For $1 \leqslant i, j \leqslant m$ we first define $p_{i,j} = (m+j-i)m-i$ and $q_{i,j} = (m+i-j+2)m+i$. Furthermore, we set $u_{i,j} = a^{p_{i,j}} b a^{q_{i,j}}$. We claim that for each i, j there exists a path from state k to ℓ labeled by $u_{i,j}$ if and only if $(k, \ell) = (i, j+m)$ or $k = \ell$.

The two a-cycles have length m and $m+1$, respectively. Since for each pair (i, j) we have $p_{i,j} + q_{i,j} = 2m(m+1)$ and since one can always stay in the same state when reading the letter b, there clearly exists a path from each state to itself labeled by $u_{i,j}$. Now, fix some (i, j) and let $(k, \ell) = (i, j+m)$. We have $i + p_{i,j} = (m+j-i)m$ which means that, when starting in state i, one can reach state m by reading $a^{p_{i,j}}$. Being in state m, one of the b-transitions leads to state $m+1$. From there on, we make a single step backwards whenever reading the factor a^m. Thus, by reading the word $a^{q_{i,j}}$, we perform $(m+i-j+2)-i = m-j+2$ backward steps in total, finally reaching state $n+1-(m-j+2) = 2m+2-(m-j+2) = m+j = \ell$. The converse direction of our claim follows immediately since the automaton is deterministic when restricted to a-transitions and since one can only reach states $\ell > m$ by using the transition $(m, b, m+1)$.

For $X \subseteq M \times M$, we now define u_X as the concatenation of all $u_{i,j}$ with $(i, j) \in X$, where the factors are ordered according to their indices (i, j). By

the above argument, it is easy to see that there is a path from state i to $j+m$ labeled by u_X if and only if $(i,j) \in X$. Since there are $2^{m^2} = 2^{(n-1)^2/4}$ subsets of the Cartesian product $M \times M$, it remains to show that for different subsets $X, Y \subseteq M \times M$, we have $u_X \not\equiv_L v_Y$. To this end, assume without loss of generality that $(i,j) \in X \setminus Y$. Then $a^{i-1}u_X a^{n-j} \in L$ but $a^{i-1}u_Y a^{n-j} \notin L$, as desired.

For the Büchi case note that for all $i \in Q$, we have $(i,b,n) \in \delta$ if and only if $i = n$. Therefore, by Proposition 4 and the arguments above, the smallest semigroup weakly recognizing $L_{\mathrm{BA}}(\mathcal{A})$ has at least $2^{(n-1)^2/4}$ elements. □

The construction above does not reach the 2^{n^2} bound obtained when using a larger alphabet. However, this is not surprising, given the following result.

Proposition 8. *Let $m \in \mathbb{N}$ be a fixed integer and let A be an alphabet of size m. Then there exists an integer $n_m \geqslant 1$ such that for each finite automaton $\mathcal{A}$ over A with $n \geqslant n_m$ states, the language $L_{NFA}(\mathcal{A}) \subseteq A^*$ (resp. $L_{BA}(\mathcal{A}) \subseteq A^\omega$) is recognized (resp. weakly recognized) by a morphism onto a semigroup with less than 2^{n^2} elements.*

We do not give a full proof of the proposition here, but the claim essentially follows from a careful analysis of the subsemigroup of the transition semigroup generated by the transitions corresponding to the letters in A. Applying Devadze's Theorem [4, 8] to the matrix representation of this subsemigroup shows that it is proper, i.e., smaller than the full transition semigroup itself.

5 From Weakly Recognizing Morphisms to Automata

The well-known construction to convert weakly recognizing morphisms to finite automata with a Büchi-acceptance condition has quadratic blow-up [10]. We show that this is optimal up to a constant factor.

Theorem 9. *Let $A = \{a,b\}$, let $n \geqslant 3$, and let $L = \bigcup_{i=1}^{n} (ba^ibA^*)^\omega$. Then there exists a semigroup with $4n+3$ elements that weakly recognizes L and every finite automaton Büchi-accepting L has at least $n(n+1)/2$ states.*

Proof. We first define a semigroup $S = \{a^i, a^ib, ba^i, ba^ib \mid 1 \leqslant i \leqslant n\} \cup \{b, bb, 0\}$ by the multiplication $0 \cdot s = s \cdot 0 = 0$ for all $s \in S$ and

$$b^\ell a^i b^r \cdot b^m a^j b^s = \begin{cases} bb & \text{if } i = j = 0 \\ b^\ell a^{i+j} b^s & \text{if } r = m = 0 \text{ and } 1 \leqslant i+j \leqslant n \\ 0 & \text{if } r = m = 0 \text{ and } i+j > n \\ b^\ell a^i b & \text{otherwise} \end{cases}$$

where $\ell, m, r, s \in \{0,1\}$ and $i, j \in \{0, \ldots, n\}$. The morphism $h\colon A^+ \to S$ defined by $h(a) = a$ and $h(b) = b$ now weakly recognizes L since L is the union of all sets $[ba^ib][ba^ib]^\omega$ with $1 \leqslant i \leqslant n$.

Now assume that we are given a finite automaton $\mathcal{A} = (Q, A, \delta, I, F)$ such that $L_{\mathrm{BA}}(\mathcal{A}) = L$. For each $i \in \{1, \ldots, n\}$, we consider the word $\alpha_i = (ba^ib)^\omega$

and let r_i be an accepting run of α_i. We first show that for $i \neq j$, we have $\inf(r_i) \cap Q \cap \inf(r_j) = \emptyset$, and then prove that $|\inf(r_i) \cap Q| \geqslant i$ for $1 \leqslant i \leqslant n$. Together, this yields

$$|Q| \geqslant \sum_{i=1}^{n} |\inf(r_i) \cap Q| \geqslant \sum_{i=1}^{n} i = n(n+1)/2.$$

Let $i, j \in \{1, \ldots, n\}$ such that $i \neq j$. We assume for the sake of contradiction that there exists a state $q \in Q$ with $q \in \inf(r_i)$ and $q \in \inf(r_j)$. Let $u \in ba^ibA^*$ be a prefix of α_i such that r_i visits q after reading u. Let $v \in A^*$ be a factor of α_j such that there exists a finite run labeled by v, which starts and ends in q, visits at least one final state and such that $v^\omega = (ba^jb)^\omega$ or $v^\omega = a^kb(ba^jb)^\omega$ for some $k \in \{0, \ldots, j\}$. Obviously, we then have $uv^\omega \in L_{\mathrm{BA}}(\mathcal{A})$ but $uv^\omega \notin L$, a contradiction.

For the second part of the proof, assume again for the sake of contradiction that $|\inf(r_i) \cap Q| < i$ for some accepting run r_i of α_i. Then inside each ba^ib-factor, a state is visited twice and we can apply the standard pumping argument to show that a word in $A^\omega \setminus L_{\mathrm{BA}}(\mathcal{A})$ has an accepting run as well. □

6 Complementation

To date, the best construction for complementing weakly recognizing morphisms is the so-called *strong expansion* [10]. Given a morphism $h\colon A^+ \to S$, the strong expansion of h is a morphism $g\colon A^+ \to T$ which strongly recognizes all languages weakly recognized by h. If S has n elements, the size of T is 2^{n^2}. The purpose of this section is to give a lower bound for complementation. At the same time, the established bound also serves as a lower bound for the conversion of weak recognition to strong recognition since each morphism strongly recognizing a language also strongly recognizes its complement.

Complementing weakly recognizing morphisms is easy in the case of $\mathcal{J}$-trivial semigroups since each language weakly recognized by a $\mathcal{J}$-trivial semigroup S is already strongly recognized by S, i.e., there is no need the compute the strong expansion if the $\mathcal{J}$-classes of the input are trivial already. In order to establish a lower bound, we thus consider the class of simple semigroups, which is dual to $\mathcal{J}$-trivial semigroups in the sense that simple semigroups consist of a single $\mathcal{J}$-class only.

Proposition 10. *Let $n \geqslant 1$ be an arbitrary integer and let $A = \{a_1, a_2, \ldots, a_n\}$. The language $L = \bigcup_{i=1}^{n} (a_iA^*)^\omega$ is weakly recognized by a simple semigroup with n elements and every semigroup weakly recognizing $A^\omega \setminus L$ has at least $n2^{n-1}$ elements.*

Proof. The alphabet A can be extended to a semigroup by defining an associative operation $a \circ b = a$ for all $a, b \in A$. Now, the morphism $h\colon A^+ \to (A, \circ)$ given by $h(a) = a$ for all $a \in A$ weakly recognizes L. The semigroup $(A, \circ)$ contains $|A| = n$ elements and it is simple because we have $a \mathrel{\mathcal{L}} b$ for all $a, b \in A$.

Now, let $h\colon A^+ \to S$ be a morphism weakly recognizing $A^\omega \setminus L$. For a letter $b \in A$ and a subset $B \subseteq A \setminus \{b\}$, let $u_{b,B}$ be the uniquely defined word $ba_{i_1}a_{i_2}\cdots a_{i_\ell}$ such that $i_1 < i_2 < \cdots < i_\ell$ and $\{a_{i_1}, a_{i_2}, \ldots, a_{i_\ell}\} = B$. Consider two letters $b, c \in A$ and subsets $B \subseteq A \setminus \{b\}$, $C \subseteq A \setminus \{c\}$. If $b \neq c$, we have $u_{b,B}c^\omega \notin L$ and $u_{c,C}c^\omega \in L$. If $B \neq C$ we may assume, without loss of generality, that there exists a letter $a \in B \setminus C$. In this case, we have $au_{c,C}{}^\omega \notin L$ but $a(u_{b,B}u_{c,C})^\omega \in L$ and $a(u_{c,C}u_{b,B})^\omega \in L$. By Theorem 3, this suffices to conclude that $h(u_{b,B}) \neq h(u_{c,C})$ whenever $b \neq c$ or $B \neq C$ and therefore, S contains at least $|A|\, 2^{|A|-1} = n2^{n-1}$ elements. □

Rather surprisingly, the established lower bound turns out to be asymptotically tight in the case of simple semigroups. More generally, for simple semigroups, the construction of the strong expansion can be improved such that only $n2^n$ elements are needed. This will be proved in the remainder of this section.

We start with a morphism $h\colon A^+ \to S$ onto a simple semigroup with $n = |S|$ elements. Since S is simple, there exists a surjective mapping $\gamma\colon S \to G$ onto a finite group G that becomes a bijection when restricted to a single $\mathcal{H}$-class. Therefore, the mapping $\pi\colon (S/\mathcal{R}) \times G \times (S/\mathcal{L}) \to S$ with $\pi^{-1}(s) = (R_s, \gamma(s), L_s)$ for all $s \in S$ is well-defined and bijective. Moreover, for $s, t \in S$, we write $R_t \cdot s$ to denote the element $\pi(R_t, \gamma(s), L_s)$.

Let $T = \{(s, X) \mid s \in S, X \subseteq S\}$ and let $g\colon A^+ \to T$ be defined by

$$g(u) = (h(u), \{R_{h(q)} \cdot h(p) \mid p, q \in A^+, pq = u\})$$

for all $u \in A^+$. The set T can be extended to a semigroup by defining an associative multiplication

$$(s, X) \cdot (t, Y) = (st, X \cup \{R_t \cdot s\} \cup \hat{Y})$$

where $\hat{Y}$ denotes the set $\{\pi(R_y, \gamma(s(R_t \cdot y)), L_y) \mid y \in Y\}$. Under this extension, the mapping g becomes a morphism.

The following three technical lemmas capture important properties of the construction and are needed for the main proof.

Lemma 11. *Let $s, t \in S$. Then $R_t \cdot s$ is the unique element x such that $x \mathrel{\mathcal{R}} t$, $x \mathrel{\mathcal{L}} s$ and $\gamma(x) = \gamma(s)$ or, equivalently, the unique element x such that $x \mathrel{\mathcal{H}} ts$ and $\gamma(x) = \gamma(s)$.*

Proof. Let $x = R_t \cdot s$. We have $(R_x, \gamma(x), L_x) = \pi^{-1}(x) = \pi^{-1}(R_t \cdot s) = (R_t, \gamma(s), L_s)$. Together with the fact that π is bijective, this establishes the first claim. For the second claim, note that since S is simple, $x \mathrel{\mathcal{R}} t$ is equivalent to $x \mathrel{\mathcal{R}} ts$ and $x \mathrel{\mathcal{L}} s$ is equivalent to $x \mathrel{\mathcal{L}} ts$. □

Lemma 12. *Let $u \in A^+$ with $g(u) = (s, X)$ and let $x \in S$. Then $x \in X \cup \{s\}$ if and only if there exists a factorization $u = pq$ with $p \in A^+$ and $q \in A^*$ such that $x \mathrel{\mathcal{H}} h(qp)$ and $\gamma(x) = \gamma(h(p))$.*

Proof. Obviously, we have $x = s$ if and only if there exists a factorization $u = pq$ with $p = u$ and $q = \varepsilon$ satisfying the properties described above. Thus, it suffices to consider factorizations where $p, q \in A^+$. By Lemma 11, such a factorization exists if and only if $x = R_{h(q)} \cdot h(p)$ which is, in turn, equivalent to $x \in X$ by the definition of g. □

Lemma 13. *Let (t, f) be a linked pair of S, let $\big((s, X), (e, Y)\big)$ be a linked pair of T and let $\alpha \in [(s, X)]_g[(e, Y)]_g^\omega$. Then $\alpha \in [t]_h[f]_h^\omega$ if and only if $tq = s$, $pq = e$, $qp = f$, $R_q \cdot t \in X$ and $R_q \cdot p \in Y$ for some $p, q \in S$.*

Proof. For the direction from left to right, let $\alpha = uv_1v_1'v_2v_2' \cdots$ such that $g(u) = (s, X)$, $g(v_iv_i') = (e, Y)$, $h(uv_1) = t$ and $h(v_i'v_{i+1}) = f$ for all $i \geqslant 1$. Furthermore, we assume without loss of generality that $v_i, v_i' \neq \varepsilon$ for all $i \geqslant 1$ and that $h(v_1) = h(v_2)$. We set $p = h(v_1) = h(v_2)$ and $q = h(v_1')$. Now, $tq = h(uv_1v_1') = se = s$, $pq = h(v_1v_1') = e$ and $qp = h(v_1'v_2) = f$. Moreover, by the definition of g, we have $R_q \cdot t = R_{h(v_1')} \cdot h(uv_1) \in X$ and $R_q \cdot p = R_{h(v_1')} \cdot h(v_1) \in Y$.

For the converse implication, note that by Lemma 12, there exists a factorization $\alpha = uv_1v_1'v_2v_2' \cdots$ such that $h(u) = s$, $h(v_iv_i') = e$, $R_{h(v_1')} \cdot h(uv_1) = R_q \cdot t$ and $R_{h(v_i')} \cdot h(v_i) = R_q \cdot p$ for all $i \geqslant 1$. Since S is simple, $h(v_i) \mathcal{R} h(v_iv_i') = e \mathcal{R} p$ and $h(v_i) \mathcal{L} (R_{h(v_i')} \cdot h(v_i)) = (R_q \cdot p) \mathcal{L} p$ for all $i \geqslant 1$. Furthermore, $\gamma(h(v_i)) = \gamma(R_{h(v_i')} \cdot h(v_i)) = \gamma(R_q \cdot p) = \gamma(p)$. Together, this yields $h(v_i) = p$ by Lemma 11. Similarly, we have $h(v_i') \mathcal{R} (R_{h(v_i')} \cdot h(v_i)) = (R_q \cdot p) \mathcal{R} q$ and thus, $ph(v_i') = h(v_iv_i') = pq$ implies $h(v_i') = q$ for all $i \geqslant 1$ by Lemma 1. This shows that $h(uv_1) = sp = tqp = tf = t$ and $h(v_i'v_{i+1}) = qp = f$. We conclude that $\alpha \in [t][f]^\omega$. □

Theorem 14. *Let $h\colon A^+ \to S$ be a morphism onto a simple semigroup of size $n = |S|$ that weakly recognizes a language $L \subseteq A^\omega$. Then there exists a morphism $g\colon A^+ \to T$ to a semigroup of size $|T| = n2^n$ that strongly recognizes L.*

Proof. The construction we use is the one described in the introduction of this section. Consider a linked pair $((s, X), (e, Y))$ of T as well as two infinite words $\alpha, \beta \in [(s, X)][(e, Y)]^\omega$. If $\alpha \in L$, there exists a linked pair (t, f) of S such that $\alpha \in [t][f]^\omega \subseteq L$. Lemma 13 immediately yields $\beta \in [t][f]^\omega \subseteq L$, thereby showing that g strongly recognizes L. □

7 Discussion and Open Problems

We presented lower bound techniques and gave tight bounds for the conversion between finite automata and weakly recognizing morphisms. One can use techniques similar to those described in Section 4 to obtain a 3^{n^2} lower bound for the conversion of finite automata with transition-based Büchi acceptance to strongly recognizing morphisms. However, with the usual state-based Büchi acceptance criterion, the analysis becomes much more involved and it is not clear whether the 3^{n^2} upper bound can be reached. Analogously, there is no straightforward

adaptation of the conversion of weakly recognizing morphisms into Büchi automata in Section 5 to strongly recognizing morphisms. It would be interesting to see whether the quadratic lower bound also holds in this setting.

Another open problem is to close the remaining gaps between the upper and the lower bounds. This is particularly true for the complexity of complementation and the conversion of weakly recognizing morphisms to strong recognition. We showed that there is an exponential lower bound and gave an asymptotically optimal construction for simple semigroups which was a first candidate for semigroups that are hard to complement. It is easy to adapt this construction to families of semigroups where the size of each $\mathcal{J}$-class is bounded by a constant. However, for the general case, the gap between $n2^{n-1}$ and 2^{n^2} remains.

Beyond that, another direction for future research is to investigate whether any of the bounds can be improved by considering the size of the accepting set, i.e., the number of linked pairs used to describe a language.

Acknowledgments. We thank the anonymous referees for several useful suggestions which helped to improve the presentation of this paper.

References

1. A. Arnold. A syntactic congruence for rational ω-languages. *Theoretical Computer Science*, 39:333–335, 1985.
2. J. R. Büchi. Weak second-order arithmetic and finite automata. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, 6:66–92, 1960.
3. M. Chrobak. Finite automata and unary languages. *Theoretical Computer Science*, 47(2):149–158, Nov. 1986.
4. H. M. Devadze. Generating sets of the semigroup of all binary relations in a finite set. *Doklady Akademii Nauk BSSR*, 12:765–768, 1968.
5. L. Fleischer and M. Kufleitner. Efficient Algorithms for Morphisms over Omega-Regular Languages. In *FSTTCS 2015, Proceedings*, volume 45 of *LIPIcs*, pages 112–124. Dagstuhl Publishing, 2015.
6. M. Holzer and B. König. On deterministic finite automata and syntactic monoid size. *Theoretical Computer Science*, 327(3):319–347, Nov. 2004.
7. K. H. Kim and F. W. Roush. Two-generator semigroups of binary relations. *Journal of Mathematical Psychology*, 17(3):236–246, 1978.
8. J. Konieczny. A proof of Devadze's theorem on generators of the semigroup of boolean matrices. *Semigroup Forum*, 83(2):281–288, 2011.
9. J. Pécuchet. Variétés de semis groupes et mots infinis. In *STACS 1986, Proceedings*, pages 180–191, 1986.
10. D. Perrin and J.-É. Pin. *Infinite words*, volume 141 of *Pure and Applied Mathematics*. Elsevier, 2004.
11. J.-É. Pin. *Varieties of Formal Languages*. North Oxford Academic, 1986.
12. W. J. Sakoda and M. Sipser. Nondeterminism and the size of two way finite automata. In *STOC 1978, Proceedings*, pages 275–286. ACM Press, 1978.
13. W. Thomas. Automata on infinite objects. In *Handbook of Theoretical Computer Science*, chapter 4, pages 133–191. Elsevier, 1990.
14. Q. Yan. Lower bounds for complementation of omega-automata via the full automata technique. *Logical Methods in Computer Science*, 4(1), 2008.