



**HAL**  
open science

# The Complexity of Languages Resulting from the Concatenation Operation

Galina Jirásková, Alexander Szabari, Juraj Šebej

► **To cite this version:**

Galina Jirásková, Alexander Szabari, Juraj Šebej. The Complexity of Languages Resulting from the Concatenation Operation. 18th International Workshop on Descriptive Complexity of Formal Systems (DCFS), Jul 2016, Bucharest, Romania. pp.153-167, 10.1007/978-3-319-41114-9\_12. hal-01633946

**HAL Id: hal-01633946**

**<https://inria.hal.science/hal-01633946v1>**

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

# The Complexity of Languages Resulting from the Concatenation Operation

Galina Jirásková<sup>1,\*</sup>, Alexander Szabari<sup>2,\*\*</sup>, and Juraj Šebej<sup>2,\*\*</sup>

<sup>1</sup> Mathematical Institute, Slovak Academy of Sciences,  
Grešákova 6, 040 01 Košice, Slovakia  
jiraskov@saske.sk

<sup>2</sup> Institute of Computer Science, Faculty of Science, P.J. Šafárik University,  
Jesenná 5, 040 01 Košice, Slovakia  
alexander.szabari@gmail.com, juraj.sebej@gmail.com

**Abstract.** We prove that for all  $m, n$ , and  $\alpha$  with  $1 \leq \alpha \leq f(m, n)$ , where  $f(m, n)$  is the state complexity of the concatenation operation, there exist a minimal  $m$ -state DFA  $A$  and a minimal  $n$ -state DFA  $B$ , both defined over an alphabet  $\Sigma$  with  $|\Sigma| \leq 2n+4$ , such that the minimal DFA for the language  $L(A)L(B)$  has exactly  $\alpha$  states. This improves a similar result in the literature that uses an exponential alphabet.

## 1 Introduction

Iwama et al. [4] stated the question of whether there always exists a minimal nondeterministic finite automaton (NFA) of  $n$  states whose equivalent minimal deterministic finite automaton (DFA) has  $\alpha$  states for all integers  $n$  and  $\alpha$  satisfying  $n \leq \alpha \leq 2^n$ . The question was also considered by Iwama et al. [5], and answered positively in [9] for a ternary alphabet. However, in the unary case, the existence of holes, so called “magic numbers”, was proved by Geffert [1]. The binary case is still open.

The same problem on sub-regular language families was studied by Holzer et al. [2]. It turned out that the existence of non-trivial magic numbers is rare, and that the ranges of possible complexities are usually contiguous. One interesting exception was obtained by Čevorová [18]. She studied the star operation on unary regular languages, and proved that there are two linear segments of magic numbers in the range from 1 to  $(n-1)^2+1$ , that is, of values that cannot be met by the state complexity of the star of a unary language accepted by a minimal  $n$ -state DFA. On the other hand, she proved that for the square operation in the unary case no magic numbers exist [19]. Another example of the existence of magic numbers for symmetric difference NFAs was presented by Zijl [17], but they could possibly be trivial.

A similar problem for the reversal, star, and concatenation operation was studied in [7, 8], where it was shown that for all the three operations the whole

---

\* Research supported by grant VEGA 2/0084/15.

\*\* Research supported by grant VEGA 1/0142/15.

range of possible complexities up to known upper bounds can be produced using an exponential alphabet.

The result for reversal and star was improved in [10, 14] by showing that a linear alphabet is enough to produce the whole range of complexities.

In this paper we complement these results, and show that a linear alphabet can also be used for the concatenation operation. We prove that for all  $m, n$ , and  $\alpha$  with  $1 \leq \alpha \leq f(m, n)$ , where  $f(m, n)$  is the state complexity of the concatenation operation, there exist a minimal  $m$ -state DFA  $A$  and a minimal  $n$ -state DFA  $B$ , both defined over an alphabet  $\Sigma$  with  $|\Sigma| \leq 2n + 4$ , such that the minimal DFA for the language  $L(A)L(B)$  has exactly  $\alpha$  states.

To get this result, we describe three constructions, in which we are able to get  $m$ -state and  $(n + 1)$ -state DFAs  $A_i, B_i$  for  $i = 1, 2, 3$  from  $m$ -state and  $n$ -state DFAs  $A$  and  $B$ , by adding a new state to  $B$ , and by adding the transitions on two new symbols. Moreover, if the state complexity of the concatenation of  $L(A)$  and  $L(B)$  is  $\alpha$ , then the state complexity of the concatenation of  $L(A_i)$  and  $L(B_i)$ ,  $i = 1, 2, 3$ , is  $2\alpha, 2\alpha - 1$ , and  $\alpha + 1$ , respectively. As a result, we get a contiguous range of complexities from  $m + n + 1$  up to known upper bound for a linear alphabet. To get complexities from 1 to  $m + n - 1$ , we use a known result from [8]. We deal with the value  $m + n$  separately, and use a binary alphabet here.

The paper is organized as follows. The next section contains some definitions and preliminary results. In Section 3, we recall known results concerning the state complexity of concatenation. In Section 4, we prove that the range of possible complexities for the languages resulting from the concatenation operation is contiguous from 1 up to known upper bound, and we show that a linear alphabet is enough for this. Section 5 contains some concluding remarks.

## 2 Preliminaries

In this section we give some basic definitions and preliminary results. For details, the reader may refer to [3, 13, 15].

Let  $\Sigma$  be a finite alphabet of symbols. Then  $\Sigma^*$  denotes the set of strings over  $\Sigma$  including the empty string  $\varepsilon$ . The length of a string  $w$  is denoted by  $|w|$ , and the number of occurrences of a symbol  $a$  in a string  $w$  is denoted by  $\#_a(w)$ . A language is any subset of  $\Sigma^*$ . The concatenation of languages  $K$  and  $L$  is the language  $KL = \{uv \mid u \in K \text{ and } v \in L\}$ . The cardinality of a finite set  $A$  is denoted by  $|A|$ , and its power-set by  $2^A$ .

A *nondeterministic finite automaton* (NFA) is a quintuple  $A = (Q, \Sigma, \cdot, I, F)$ , where  $Q$  is a finite set of states,  $\Sigma$  is a finite alphabet,  $\cdot : Q \times \Sigma \rightarrow 2^Q$  is the transition function which is extended to the domain  $2^Q \times \Sigma^*$  in the natural way,  $I \subseteq Q$  is the set of initial states, and  $F \subseteq Q$  is the set of final states. The *language accepted by*  $A$  is the set  $L(A) = \{w \in \Sigma^* \mid I \cdot w \cap F \neq \emptyset\}$ . For a symbol  $a$ , we say that  $(p, a, q)$  is a transition in NFA  $A$  if  $q \in p \cdot a$ , and for a string  $w$ , we write  $p \xrightarrow{w} q$  if  $q \in p \cdot w$ . We say that  $(p, a, q)$  is an in-transition going to state  $q$ .

An NFA  $A$  is *deterministic* (DFA) (and complete) if  $|I| = 1$  and  $|q \cdot a| = 1$  for each  $q$  in  $Q$  and each  $a$  in  $\Sigma$ . In such a case, we write  $q \cdot a = q'$  instead of  $q \cdot a = \{q'\}$ .

The *state complexity* of a regular language  $L$ ,  $\text{sc}(L)$ , is the smallest number of states in any DFA for  $L$ . The state complexity of a binary regular operation  $\circ$  is defined as a function  $f(m, n)$  given by

$$f(m, n) = \max\{\text{sc}(K \circ L) \mid K, L \subseteq \Sigma^*, \text{sc}(K) = m, \text{sc}(L) = n\}.$$

Every NFA  $A = (Q, \Sigma, \cdot, I, F)$  can be converted to an equivalent DFA  $A' = (2^Q, \Sigma, \cdot', I, F')$ , where  $R \cdot' a = R \cdot a$  and  $F' = \{R \in 2^Q \mid R \cap F \neq \emptyset\}$  [12]. The DFA  $A'$  is called the *subset automaton* of the NFA  $A$ . The subset automaton may not be minimal since some of its states may be unreachable or equivalent to other states.

In the following proposition, we provide a sufficient condition for an NFA, which guarantees that the corresponding subset automaton does not have equivalent states.

**Proposition 1.** *Let  $N = (Q, \Sigma, \cdot, I, F)$  be an NFA. Assume that for each state  $q$  in  $Q$ , there is a string  $w_q$  in  $\Sigma^*$  which is accepted by  $N$  only from the state  $q$ , that is, we have  $q \cdot w_q \cap F \neq \emptyset$ , and  $p \cdot w_q \cap F = \emptyset$  if  $p \neq q$ . Then the subset automaton of  $N$  does not have equivalent states.*

*Proof.* Let  $S$  and  $T$  be two distinct subsets of the subset automaton. Then, without loss of generality, there is a state  $q$  with  $q \in S \setminus T$ . Then the string  $w_q$  is accepted by the subset automaton from the subset  $S$ , but it is rejected from  $T$ .  $\square$

To describe string  $w_q$  accepted by an NFA only from state  $q$ , we usually use the next observation.

**Proposition 2.** *Let a string  $w_q$  be accepted by an NFA  $N$  only from state  $q$ . If  $(p, a, q)$  is the unique in-transition going to state  $q$  by symbol  $a$ , then the string  $aw_q$  is accepted by  $N$  only from state  $p$ .*

In what follows, we often need to show how the set of all the reachable subsets in a subset automaton looks like. To do this, the following observation is useful.

**Proposition 3.** *Let  $D$  be a subset automaton of an NFA  $N = (Q, \Sigma, \cdot, I, F)$ . Let  $\mathcal{R}$  be a family of subsets of  $Q$  such that*

- (1) *each subset in  $\mathcal{R}$  is reachable in  $D$ ,*
- (2)  *$I \in \mathcal{R}$ , and*
- (3) *for each  $S$  in  $\mathcal{R}$  and each symbol  $a$  in  $\Sigma$ , the set  $S \cdot a$  is in  $\mathcal{R}$ .*

*Then  $\mathcal{R}$  is the family of all reachable subsets of DFA  $D$ .*

*Proof.* Each set in  $\mathcal{R}$  is reachable in  $D$  by (1). Let  $S$  be a reachable subset of  $D$ . Then there is a string  $w$  in  $\Sigma^*$  such that  $S = I \cdot w$ . We prove the proposition by induction on  $|w|$ . If  $|w| = 0$ , then  $w = \varepsilon$  and  $S = I \cdot \varepsilon = I$ , which is in  $\mathcal{R}$  by (2). Now let  $w = va$  for a string  $v$  and a symbol  $a$ . By the induction hypothesis, the set  $S' = I \cdot v$  is in  $\mathcal{R}$ . Then  $S = S' \cdot a$ , so  $S$  is in  $\mathcal{R}$  by (3).  $\square$

### 3 State Complexity of Concatenation

Consider minimal DFAs  $A$  and  $B$ . Without loss of generality, we assume that the state set of  $A$  is  $\{q_0, q_1, \dots, q_{m-1}\}$  with the initial state  $q_0$ , and the state set of  $B$  is  $\{0, 1, \dots, n-1\}$  with the initial states  $0$ . Moreover, in both  $A$  and  $B$ , let us denote the transition function by  $\cdot$ . This is not confusing since the state sets of  $A$  and  $B$  are disjoint. First, let us recall the construction of an NFA for the language  $L(A)L(B)$ .

**Construction of NFA for concatenation:**

(DFA  $A$  and DFA  $B \rightarrow$  NFA  $N$  for  $L(A)L(B)$ )

Let  $A = (\{q_0, q_1, \dots, q_{m-1}\}, \Sigma, \cdot, q_0, F_A)$  and  $B = (\{0, 1, \dots, n-1\}, \Sigma, \cdot, 0, F_B)$  be DFAs. Construct NFA  $N$  from DFAs  $A$  and  $B$  as follows:

- (a) for each symbol  $a$  and each state  $q_i$  with  $q_i \cdot a \in F_A$ , add transition  $(q_i, a, 0)$ ;
- (b) the set of initial states of  $N$  is  $\{q_0\}$  if  $q_0 \notin F_A$ , and it is  $\{q_0, 0\}$  otherwise;
- (c) the set of final state of  $N$  is  $F_B$ .

In the subset automaton of NFA  $N$  constructed as above, each reachable subset is of the form  $\{q_i\} \cup S$ , where  $S \subseteq \{0, 1, \dots, n-1\}$  since  $A$  is deterministic and complete. Moreover, if  $q_i$  is a final state of  $A$ , then  $0 \in S$  since  $N$  has the transition  $(q, a, 0)$  whenever a state  $q$  of  $A$  goes to a final state  $q_i$  on a symbol  $a$ . It follows that the subset automaton of  $N$  has at most  $(m-k)2^n + k2^{n-1}$  reachable states. Next we have  $(m-k)2^n + k2^{n-1} = m2^n - k2^{n-1}$ , which is maximal if  $k = 1$  [11, 16]. We write this upper bound as  $(m-1)2^n + 2^{n-1}$ . The bound is known to be tight if  $m \geq 1$  and  $n \geq 2$  [6, 11, 16]. If  $m \geq 1$  and  $n = 1$ , then  $L = \emptyset$  or  $L = \Sigma^*$ , so the tight upper bound in this case is  $m$ . Hence we get the following result.

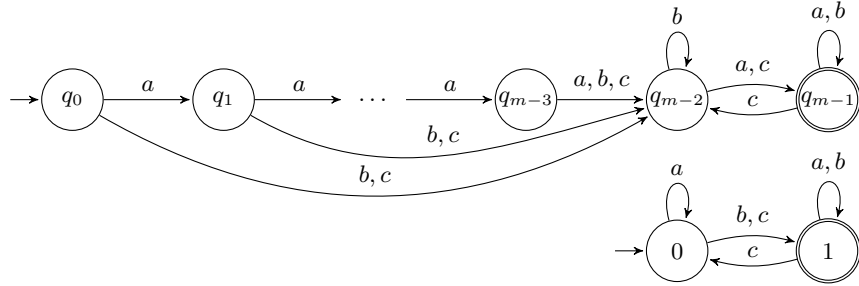
**Proposition 4 ([6, 11, 16]).** *Let  $m, n \geq 1$  and  $f(m, n)$  be the state complexity of the concatenation operation on languages over an alphabet of size at least two defined as  $f(m, n) = \max\{\text{sc}(KL) \mid K, L \subseteq \Sigma^*, |\Sigma| \geq 2, \text{sc}(K) = m, \text{sc}(L) = n\}$ . Then we have*

$$f(m, n) = \begin{cases} m, & \text{if } n = 1; \\ (m-1)2^n + 2^{n-1}, & \text{if } n \geq 2. \end{cases}$$

### 4 The Range of Possible Complexities

The aim of this section is to show that the whole range of complexities from 1 to  $f(m, n)$  for the concatenation operation can be produced using an alphabet that grows linearly with  $n$ .

To this aim consider minimal DFAs  $A = (\{q_0, q_1, \dots, q_{m-1}\}, \Sigma, \cdot, q_0, \{q_{m-1}\})$ , and  $B = (\{0, 1, \dots, n-1\}, \Sigma, \cdot, 0, \{1\})$ . Construct an NFA  $N$  for  $L(A)L(B)$  as described in Section 3. Let  $D$  be the subset automaton of  $N$ , and  $\mathcal{R}$  the family of all the reachable subsets in DFA  $D$ . We assume that  $A, B, N, D$ , and  $\mathcal{R}$  satisfy the following conditions.



**Fig. 1.** Transitions on  $a, b, c$  in states in  $\{q_0, q_1, \dots, q_{m-1}\} \cup \{0, 1\}$ .

- (1) The transitions on symbols  $a, b, c$  in states in  $\{q_0, q_1, \dots, q_{m-1}\} \cup \{0, 1\}$  are defined as in Fig. 1.
- (2) If  $(q_i, \sigma, q_0)$  is a transition in  $A$  for some  $\sigma$  in  $\Sigma$ , then  $i = m - 1$ .
- (3) Each set in  $\mathcal{R} \setminus \{\{q_0\}\}$  is reachable from  $\{q_1\}$  in the subset automaton  $D$ .
- (4) For each state  $q$  of NFA  $N$ , there exists a string  $w_q$  in  $\Sigma^*$  accepted by  $N$  only from state  $q$ . Moreover, we have

$$\begin{aligned}
 w_1 &= \varepsilon, \\
 w_0 &= c, \\
 w_{q_{m-1}} &= bc, \\
 w_{q_{m-2}} &= cbc, \\
 w_{q_{m-2-i}} &= a^i cbc \text{ for } i = 1, 2, \dots, m - 2, \text{ and} \\
 w_j &= a_j \text{ for } j = 2, 3, \dots, n - 1.
 \end{aligned}$$

**Proposition 5.** Let  $A, B, N, D$ , and  $\mathcal{R}$  satisfy conditions (1)–(4). Then

- (a) The sets  $\{q_1\}, \{q_{m-1}, 0\}, \{q_{m-1}, 0, 1\}, \{q_{m-2}, 0, 1\}$  are in  $\mathcal{R}$ .
- (b) The initial subset  $\{q_0\}$  of the subset automaton  $D$  cannot be reached from any other reachable subset of  $D$ .
- (c) The subset automaton  $D$  of NFA  $N$  does not have equivalent states, so  $\text{sc}(L(A)L(B)) = |\mathcal{R}|$ .

*Proof.* (a) By (1), the transitions on  $a, b, c$  are as in Fig. 1. It follows that in the subset automaton  $D$ , we have

$$\{q_0\} \xrightarrow{a} \{q_1\} \xrightarrow{a^{m-2}} \{q_{m-1}, 0\} \xrightarrow{b} \{q_{m-1}, 0, 1\} \xrightarrow{c} \{q_{m-2}, 0, 1\}.$$

(b) Assume for a contradiction that there is a set  $S$  in  $\mathcal{R}$  and a symbol  $\sigma$  such that  $S \cdot \sigma = \{q_0\}$ . Then we must have  $q_{m-1} \in S$  by (2). It follows that the initial state 0 of  $B$  must be in  $S$  since  $q_{m-1}$  is final in  $A$ . However then  $S \cdot \sigma \supseteq \{q_0, 0 \cdot \sigma\}$ , a contradiction.

(c) By (4), the NFA  $N$  satisfies the condition in Proposition 1. Therefore the subset automaton  $D$  of  $N$  does not have equivalent states, and we have  $\text{sc}(L(A)L(B)) = |\mathcal{R}|$ .  $\square$

Now our goal is to construct a minimal  $m$ -state DFA  $A_i$  and a minimal  $(n+1)$ -state DFA  $B_i$  for  $i = 1, 2, 3$  over the alphabet  $\Sigma \cup \{a_n, b_n\}$  from automata  $A$  and  $B$ , such that  $A, B, N, D, \mathcal{R}$  satisfy conditions (1)–(4), in such a way that  $A_i$  and  $B_i$ , the NFA  $N_i$  for  $L(A_i)L(B_i)$ , the subset automaton  $D_i$  of  $N_i$  and the family  $\mathcal{R}_i$  of reachable states of  $D_i$  satisfy conditions (1)–(4). Moreover, if  $\mathcal{R} = \alpha$ , then  $|\mathcal{R}_1| = 2\alpha$ ,  $|\mathcal{R}_2| = 2\alpha - 1$ , and  $|\mathcal{R}_3| = \alpha + 1$ .

We construct automata  $A_i$  and  $B_i$  from automata  $A$  and  $B$  by adding a new state  $n$  to DFA  $B$ , and by adding the transitions on two new symbols  $a_n$  and  $b_n$ . The transitions on  $a_n$  are the same in all the three constructions, and they guarantee that the string  $a_n$  is accepted by  $N_i$  only from state  $n$ . The transitions on  $b_n$  are used to reach the set  $\{q_0, n\}$  in  $D_1$ , the set  $\{q_1, n\}$  in  $D_2$  and the set  $\{q_{m-1}, 0, n\}$  in  $D_3$ . We have to be careful with condition (4), especially in the third construction.

	C1	C2	C3
$\sigma \in \Sigma$	$n \rightarrow n$	$n \rightarrow n$	$n \xrightarrow{c} 0$ $n \xrightarrow{\sigma} 0 \cdot \sigma$ if $\sigma \neq c$
$a_n$	$q_i \rightarrow q_{m-1}$ $n \rightarrow 1$ $j \rightarrow 0$	$q_i \rightarrow q_{m-1}$ $n \rightarrow 1$ $j \rightarrow 0$	$q_i \rightarrow q_{m-1}$ $n \rightarrow 1$ $j \rightarrow 0$
$b_n$	$q_{m-1} \rightarrow q_0$ $q_i \rightarrow q_{m-1}$ if $i \neq m-1$ $n \rightarrow n$ $j \rightarrow n$	$q_{m-1} \rightarrow q_1$ $q_i \rightarrow q_{m-1}$ if $i \neq m-1$ $n \rightarrow n$ $j \rightarrow n$	$q_i \rightarrow q_{m-1}$ $n \rightarrow n$ $j \rightarrow n$

**Table 1.** New transitions;  $i \in \{0, 1, \dots, m-1\}$ ,  $j \in \{0, 1, \dots, n-1\}$ .

**Construction 1.** ( $\alpha \rightarrow 2\alpha$ )

Construct DFAs  $A_1$  and  $B_1$  from DFAs  $A$  and  $B$  as follows:

- (1) add a new state  $n$  to DFA  $B$  going to itself on each old symbol  $\sigma$  in  $\Sigma$ ;
- (2) add the transitions on two new symbols  $a_n$  and  $b_n$  as shown in Table 1 in column C1.

**Construction 2.** ( $\alpha \rightarrow 2\alpha - 1$ )

Construct DFAs  $A_2$  and  $B_2$  from DFAs  $A$  and  $B$  as follows:

- (1) add a new state  $n$  to DFA  $B$  going to itself on each old symbol  $\sigma$  in  $\Sigma$ ;
- (2) add the transitions on two new symbols  $a_n$  and  $b_n$  as shown in Table 1 in column C2.

**Construction 3.** ( $\alpha \rightarrow \alpha + 1$ )

Construct DFAs  $A_3$  and  $B_3$  from DFAs  $A$  and  $B$  as follows:

- (1) add a new state  $n$  to DFA  $B$  with  $n \cdot c = 1$  and  $n \cdot \sigma = 0 \cdot \sigma$  if  $\sigma \in \Sigma \setminus \{c\}$ ;
- (2) add the transitions on two new symbols  $a_n$  and  $b_n$  as shown in Table 1 in column C3.

**Lemma 6.** *Let  $A, B, N, D, \mathcal{R}$  satisfy conditions (1)–(4). Let  $A_i, B_i$  for  $i = 1, 2, 3$  be the DFAs resulting from Constructions 1, 2, 3, respectively. Let  $N_i$  be an NFA for  $L(A_i)L(B_i)$  constructed as described in Section 3,  $D_i$  be the corresponding subset automaton, and  $\mathcal{R}_i$  be the family of all the reachable subsets in DFA  $D_i$ . Then all these automata satisfy conditions (1)–(4). Moreover, if  $|\mathcal{R}| = \alpha$ , then  $|\mathcal{R}_1| = 2\alpha$ ,  $|\mathcal{R}_2| = 2\alpha - 1$ , and  $|\mathcal{R}_3| = \alpha + 1$ .*

*Proof.* Since we do not change transitions on symbols in  $\Sigma$  on states of  $A$  and  $B$ , condition (1) is satisfied. Since the only new transition to  $q_0$  is  $(q_{m-1}, b_n, q_0)$  in Construction 1, condition (2) is satisfied in each  $A_i$ .

In each  $N_i$ , the string  $a_n$  is accepted only from state  $n$ . Moreover, in  $B_1$  and  $B_2$ , state  $n$  goes to itself on each symbol in  $\Sigma$ . It follows that condition (4) is satisfied for  $N_1$  and  $N_2$ . In  $B_3$ , we have  $n \cdot c = 0$  and  $n \cdot b = 0 \cdot b = b$ . It follows that  $(0, c, 1)$  is the only transition on  $c$  going to state 1, and  $(q_{m-1}, b, 0)$  is the only transition on  $b$  going to state 0. It follows that (4) is satisfied for  $N_3$  as well.

Now consider the subset automata  $D_1, D_2, D_3$ . Since we did not change transitions on symbols in  $\Sigma$  on states in  $A$  and  $B$ , we have  $\mathcal{R} \subseteq \mathcal{R}_i$  for  $i = 1, 2, 3$ . Let us show that

$$\begin{aligned}\mathcal{R}_1 &= \mathcal{R} \cup \{S \cup \{n\} \mid S \in \mathcal{R}\}, \\ \mathcal{R}_2 &= \mathcal{R} \cup \{S \cup \{n\} \mid S \in \mathcal{R} \text{ and } S \neq \{q_0\}\}, \\ \mathcal{R}_3 &= \mathcal{R} \cup \{\{q_{m-1}, 0, n\}\}.\end{aligned}$$

If  $S$  is in  $\mathcal{R}$  then  $S$  is reachable in  $D$ , so  $S$  can be reached from the initial state  $\{q_0\}$  by a string  $u_S$  over  $\Sigma$ . If moreover,  $S \neq \{q_0\}$ , then, by (3),  $S$  is reached from  $\{q_1\}$  by a string  $v_S$ .

In  $D_1$  we have  $\{q_0\} \xrightarrow{a} \{q_1\} \xrightarrow{a^{m-2}} \{q_{m-1}, 0\} \xrightarrow{b_n} \{q_0, n\} \xrightarrow{u_S} S \cup \{n\}$ . Thus  $\mathcal{R} \cup \{S \cup \{n\} \mid S \in \mathcal{R}\} \subseteq \mathcal{R}_1$ , and every new set  $S \cup \{n\}$  can be reached from  $\{q_1\}$ . Let us show that no other set is reachable in  $D_1$ . For each set  $S$  in  $\mathcal{R}$  and each  $\sigma$  in  $\Sigma$ , we have

$$\begin{aligned}S \cdot \sigma &\in \mathcal{R}, \\ S \cdot a_n &\in \{\{q_{m-1}, 0\}, \{q_{m-1}, 0, 1\}\}, \\ S \cdot b_n &\in \{\{q_0, n\}, \{q_{m-1}, 0\}, \{q_{m-1}, 0, n\}\}, \\ (S \cup \{n\}) \cdot \sigma &= S \cdot \sigma \cup \{n\}, \\ (S \cup \{n\}) \cdot a_n &= \{q_{m-1}, 0, 1\}, \text{ and} \\ (S \cup \{n\}) \cdot b_n &\in \{\{q_0, n\}, \{q_{m-1}, 0, n\}\}.\end{aligned}$$

Using Proposition 5(a), we get that all the resulting sets are in  $\mathcal{R} \cup \{S \cup \{n\} \mid S \in \mathcal{R}\}$ . By Proposition 3, we have  $\mathcal{R}_1 = \mathcal{R} \cup \{S \cup \{n\} \mid S \in \mathcal{R}\}$ , and, moreover,  $\mathcal{R}_1$  satisfies condition (3).

Next, in  $D_2$  we have  $\{q_0\} \xrightarrow{a} \{q_1\} \xrightarrow{a^{m-2}} \{q_{m-1}, 0\} \xrightarrow{b_n} \{q_1, n\} \xrightarrow{v_S} S \cup \{n\}$  if  $S \neq \{q_0\}$ . So every new set  $S \cup \{n\}$  is reached from  $\{q_1\}$ . The transitions on each  $\sigma$  in  $\Sigma$  and on  $a_n$  are the same as in Construction 1, and for each  $S$  in  $\mathcal{R}$ ,

$$\begin{aligned}S \cdot b_n &\in \{\{q_1, n\}, \{q_{m-1}, 0\}, \{q_{m-1}, 0, n\}\}, \text{ and} \\ (S \cup \{n\}) \cdot b_n &\in \{\{q_1, n\}, \{q_{m-1}, 0, n\}\}.\end{aligned}$$

All the resulting sets are in  $\mathcal{R} \cup \{S \cup \{n\} \mid S \in \mathcal{R} \text{ and } S \neq \{q_0\}\}$ . Moreover,  $\{q_0\}$  cannot be reached from any other subset in  $\mathcal{R}$ . By Proposition 3, we have  $\mathcal{R}_2 = \mathcal{R} \cup \{S \cup \{n\} \mid S \in \mathcal{R} \text{ and } S \neq \{q_0\}\}$ . Moreover,  $\mathcal{R}_2$  satisfies condition (3).



Finally, in  $D_3$  we have  $\{q_0\} \xrightarrow{a} \{q_1\} \xrightarrow{a^{m-2}} \{q_{m-1}, 0\} \xrightarrow{b_n} \{q_{m-1}, 0, n\}$ , so the new set  $\{q_{m-1}, 0, n\}$  is reached from  $\{q_1\}$ . The transitions on  $a_n$  are the same as above, and for each  $S$  in  $\mathcal{R}$  and each  $\sigma$  in  $\Sigma$ , we have  $S \cdot b_n \in \{\{q_{m-1}, 0\}, \{q_{m-1}, 0, n\}\}$ . Next, for the new set  $\{q_{m-1}, 0, n\}$ , we have

$$\begin{aligned} \{q_{m-1}, 0, n\} \cdot c &= \{q_{m-2}, 0, 1\}, \\ \{q_{m-1}, 0, n\} \cdot \sigma &= \{q_{m-1}, 0\} \cdot \sigma \text{ if } \sigma \in \Sigma \text{ and } \sigma \neq c; \\ \{q_{m-1}, 0, n\} \cdot b_n &= \{q_{m-1}, 0, n\}. \end{aligned}$$

All the resulting subsets are in  $\mathcal{R} \cup \{\{q_{m-1}, 0, n\}\}$ . By Proposition 3, we have  $\mathcal{R}_3 = \{\mathcal{R} \cup \{q_{m-1}, 0, n\}\}$ , and again,  $\mathcal{R}_3$  satisfies condition (3).  $\square$

Recall that  $f(m, n) = (m-1)2^n + 2^{n-1}$  is the state complexity of concatenation if  $n \geq 2$ . Our first aim is to show that each value in the range from  $m+n+1$  to  $f(m, n)$  may be attained by the state complexity of concatenation of  $m$ -state and  $n$ -state DFA languages provided that  $m \geq 3$ . We show this by induction, with the basis proved in the next lemma.

**Lemma 7.** *Let  $m \geq 3$  and  $n = 2$ . For each  $\alpha$  with  $m+3 \leq \alpha \leq f(m, 2) = 4m-2$ , there exist a minimal  $m$ -state DFA  $A$  and a minimal 2-state DFA  $B$ , both defined over an alphabet  $\Sigma$  with  $|\Sigma| \leq 7$ , such that  $\text{sc}(L(A)L(B)) = \alpha$ . Moreover, the corresponding NFA  $N$  for  $L(A)L(B)$ , the subset automaton  $D$  of  $N$ , and the set  $\mathcal{R}$  of reachable states of  $D$  satisfy conditions (1)–(4) on page 5.*

*Proof.* We first consider the values  $\alpha = i(m-2) + 6$  for  $i = 1, 2, 3, 4$ . Then we consider all the intermediate values of  $\alpha$ . Finally we deal with the case  $\alpha = m+3$ .

First let  $i = 1$ , so  $\alpha = (m-2) + 6 = m+4$ . Define a minimal  $m$ -state DFA  $A_{1,0} = (\{q_0, q_1, \dots, q_{m-1}\}, \{a, b, c, d\}, \cdot, q_0, \{q_{m-1}\})$  where for each  $i$  in  $\{0, 1, \dots, m-1\}$ ,

$$\begin{aligned} q_i \cdot a &= q_{i+1} \text{ if } i \neq m-1 \text{ and } q_{m-1} \cdot a = q_{m-1}, \\ q_i \cdot b &= q_{m-2} \text{ if } i \neq m-1 \text{ and } q_{m-1} \cdot b = q_{m-1}, \\ q_i \cdot c &= q_{m-2} \text{ if } i \neq m-2 \text{ and } q_{m-2} \cdot c = q_{m-1}, \text{ and} \\ q_i \cdot d &= q_{m-2}. \end{aligned}$$

Define a minimal two-state DFA  $B_{1,0} = (\{0, 1\}, \{a, b, c, d\}, \cdot, 0, \{1\})$  where

$$\begin{aligned} 0 \cdot a &= 0, \text{ and } 1 \cdot a = 1, \\ 0 \cdot b &= 1, \text{ and } 1 \cdot b = 1, \\ 0 \cdot c &= 1, \text{ and } 1 \cdot c = 0, \\ 0 \cdot d &= 0, \text{ and } 1 \cdot d = 1. \end{aligned}$$

Construct NFA  $N_{1,0}$  for  $L(A_{1,0})L(B_{1,0})$ , and let  $D_{1,0}$  be the corresponding subset automaton. Notice that (1), (2), and (4) are satisfied. Next, in  $D_{1,0}$  we have

$$\begin{aligned} \{q_0\} &\xrightarrow{a} \{q_1\} \xrightarrow{a^{i-1}} \{q_i\} \text{ for } i = 0, 1, \dots, m-2, \\ \{q_{m-2}\} &\xrightarrow{a} \{q_{m-1}, 0\} \xrightarrow{b} \{q_{m-1}, 0, 1\} \xrightarrow{c} \{q_{m-2}, 0, 1\}, \\ \{q_{m-1}, 0\} &\xrightarrow{d} \{q_{m-2}, 0\} \xrightarrow{c} \{q_{m-2}, 1\}. \end{aligned}$$

Thus the subset automaton has  $m+4$  reachable subsets. Next, notice that each of these  $m+4$  subsets goes to some of them by each symbol in  $\{a, b, c, d\}$ . By Proposition 3, no other set is reachable, so the complexity of  $L(A_{1,0})L(B_{1,0})$  is  $m+4$ . Notice that all the possible subsets containing states  $q_{m-1}$  and  $q_{m-2}$  are reachable in  $D_{1,0}$ .

Now we construct appropriate DFAs from automata  $A_{1,0}$  and  $B_{1,0}$  by adding transitions on new symbols. Thus we do not change the transitions on symbols  $a, b, c, d$ , and therefore the conditions (1) and (4) are always satisfied. Moreover, for each new symbol, the new transition is defined in such a way that condition (2) is satisfied as well. Finally, notice that  $\{q_{m-1}, 0\}$  is reachable from  $\{q_1\}$  by  $a^{m-2}$  in the subset automaton  $D_{1,0}$ . In what follows, we always reach new subsets in the corresponding subset automata for concatenation from the subset  $\{q_{m-1}, 0\}$ . Hence condition (3) is always satisfied.

Next, let  $\alpha = 2(m-2) + 6$ . Construct DFAs  $A_{2,0}, B_{2,0}$  from DFAs  $A_{1,0}, B_{1,0}$  by adding the transitions on a new symbol  $e_0$  as follows:

$$\begin{aligned} q_{m-1} \cdot e_0 &= q_0 \text{ and } q_i \cdot e_0 = q_{m-1} \text{ for } i = 0, 1, \dots, m-2; \\ 0 \cdot e_0 &= 0 \text{ and } 1 \cdot e_0 = 0. \end{aligned}$$

Construct the NFA  $N_{2,0}$  for  $L(A_{2,0})L(B_{2,0})$ . In the subset automaton  $D_{2,0}$ , all the sets that were reachable in the subset automaton  $D_{1,0}$  are reachable as well, since the transitions on the old symbols  $a, b, c, d$  are the same. For the same reason, the NFA  $N_{2,0}$  satisfies (4), and therefore the subset automaton  $D_{2,0}$  does not have equivalent states. Next, in  $D_{2,0}$ , we have

$$\{q_{m-1}, 0\} \xrightarrow{e_0} \{q_0, 0\} \xrightarrow{a^i} \{q_i, 0\} \text{ for } i = 1, 2, \dots, m-3.$$

No other new set is reachable since each set  $\{q_i, 0\}$  goes either to a set  $\{q_j, 0\}$  or to a set containing  $q_{m-2}$  or  $q_{m-1}$  by each symbol in  $\{a, b, c, d, e_0\}$ , and moreover, by  $e_0$ , each set goes either to  $\{q_0, 0\}$  or to a set containing  $q_{m-1}$ . Therefore the resulting complexity of the concatenation  $L(A_{2,0})L(B_{2,0})$  is  $2(m-2) + 6$ .

In a similar way, we construct DFAs  $A_{3,0}, B_{3,0}$  from  $A_{2,0}, B_{2,0}$  by adding transitions on a new symbol  $e_{01}$  defined as follows:

$$\begin{aligned} q_{m-1} \cdot e_{01} &= q_0 \text{ and } q_i \cdot e_{01} = q_{m-1} \text{ for } i = 0, 1, \dots, m-2; \\ 0 \cdot e_{01} &= 0 \text{ and } 1 \cdot e_{01} = 1. \end{aligned}$$

This results in the reachability of  $m-2$  new subsets  $\{q_i, 0, 1\}$  in the subset automaton of  $N_{3,0}$ . Since no other new set is reachable, the complexity of  $L(A_{3,0})L(B_{3,0})$  is  $3(m-2) + 6$ .

Finally, construct DFAs  $A_{4,0}, B_{4,0}$  from  $A_{3,0}, B_{3,0}$  by adding the transitions on a new symbol  $e_1$  defined as

$$\begin{aligned} q_{m-1} \cdot e_1 &= q_0 \text{ and } q_i \cdot e_1 = q_{m-1} \text{ for } i = 0, 1, \dots, m-2; \\ 0 \cdot e_1 &= 1 \text{ and } 1 \cdot e_1 = 1. \end{aligned}$$

This results in the reachability of subsets  $\{q_i, 1\}$  in the subset automaton of  $N_{4,0}$ , and the complexity of  $L(A_{4,0})L(B_{4,0})$  is  $4(m-2) + 6$ .

Up to now we have defined appropriate automata  $A_{i,0}$  and  $B_{i,0}$  for the values  $\alpha = i(m-2) + 6$  for  $i = 1, 2, 3, 4$ . Now let us consider an intermediate value  $\alpha = i(m-2) + 6 + j$  where  $1 \leq i \leq 3$  and  $1 \leq j \leq m-3$ . Construct DFAs  $A_{i,j}$  and  $B_{i,j}$  from automata  $A_{i,0}$  and  $B_{i,0}$  by adding the transitions on a new symbol  $f_1$  as follows:

$$\begin{aligned} q_{m-1} \cdot f_1 &= q_{m-2-j} \text{ and } q_i \cdot f_1 = q_{m-1} \text{ for } i = 0, 1, \dots, m-2; \\ 0 \cdot f_1 &= 1 \text{ and } 1 \cdot f_1 = 1. \end{aligned}$$

This results in the reachability of the following  $j$  new subsets in the subset

automaton of  $N_{i,j}$ :

$$\{q_{m-1}, 0\} \xrightarrow{f_1} \{q_{m-2-j}, 1\} \xrightarrow{a} \{q_{m-2-j+1}, 1\} \xrightarrow{a} \cdots \xrightarrow{a} \{q_{m-3}, 1\}.$$

Recall that the subset automaton of  $N_{i,0}$  has  $i(m-2) + 6$  reachable states, and since  $i \leq 3$ , the subsets  $\{q_i, 1\}$  are unreachable in the subset automaton of  $N_{i,0}$ . Hence the resulting complexity of  $L(A_{i,j})L(B_{i,j})$  is  $i(m-2) + 6 + j$  as desired. Moreover, all the automata satisfy conditions (1)–(4).

Finally notice that if  $A$  and  $B$  are DFAs over  $a, b, c$  shown in Fig. 1 then  $\text{sc}(L(A)L(B)) = m + 3$ . This concludes our proof.  $\square$

Now we are ready to prove the main lemma. Recall that the state complexity of concatenation is  $f(m, n) = (m-1)2^n + 2^{n-1}$  if  $n \geq 2$ . Moreover, notice that we have  $f(m, n+1) = (m-1)2^{n+1} + 2^n = 2((m-1)2^n + 2^{n-1}) = 2f(m, n)$ .

**Lemma 8.** *Let  $m \geq 3$  and  $n \geq 2$ . For each  $\alpha$  with  $m + n + 1 \leq \alpha \leq f(m, n)$ , there exist a minimal  $m$ -state DFA  $A$ , and a minimal  $n$ -state DFA  $B$ , both defined over an alphabet  $\Sigma$  with  $|\Sigma| \leq 2n + 4$ , such that  $\text{sc}(L(A)L(B)) = \alpha$ .*

*Proof.* We prove the claim by induction on  $n$ . Moreover, in the induction hypothesis, we assume that DFAs  $A$  and  $B$ , the corresponding NFA  $N$  for  $L(A)L(B)$  constructed as in Section 3, the subset automaton  $D$  of  $N$ , and the set  $\mathcal{R}$  of reachable states of  $D$  satisfy conditions (1)–(4) on page 5.

The basis, in which we have  $m \geq 3$ ,  $n = 2$ , and  $m+3 \leq \alpha \leq f(m, 2) = 4m-2$ , is proved in Lemma 7. Let  $m \geq 3$ ,  $n \geq 2$ , and assume that for each  $\beta$  with  $m + n + 1 \leq \beta \leq f(m, n)$ , there exist a minimal  $m$ -state DFA  $A$  and a minimal  $n$ -state DFA  $B$ , both defined over an alphabet  $\Sigma$  with  $|\Sigma| \leq 2n + 4$ , such that  $\text{sc}(L(A)L(B)) = \beta$ . Moreover, assume that DFAs  $A$  and  $B$ , the NFA  $N$  for  $L(A)L(B)$ , the subset automaton  $D$  of  $N$ , and the set of reachable states  $\mathcal{R}$  of  $D$  satisfy conditions (1)–(4) on page 5. Let us show that the claim holds for  $n+1$ . To this aim let  $\alpha$  be an integer with  $m + (n+1) + 1 \leq \alpha \leq f(m, n+1)$ .

First, let  $2m + 2n + 2 \leq \alpha \leq f(m, n+1)$  and  $\alpha$  be even. Let  $\beta = \alpha/2$ . Then  $m+n+1 \leq \beta \leq f(m, n)$ , and by the induction hypothesis, there exists a minimal  $m$ -state DFA  $A$  and a minimal  $n$ -state DFA  $B$ , both defined over an alphabet  $\Sigma$  with  $|\Sigma| \leq 2n + 4$ , such that  $\text{sc}(L(A)L(B)) = \beta$ . Moreover, conditions (1)–(4) are satisfied for  $A, B, N, D, \mathcal{R}$ . We use Construction 1, in which we add a new state to DFA  $B$  and the transitions on two new symbols to get a minimal  $m$ -state  $A_1$  and a minimal  $(n+1)$ -state DFA  $B_1$ . By Lemma 6, all conditions (1)–(4) are satisfied for  $A_1, B_1, N_1, D_1$ , and  $\mathcal{R}_1$ . It follows that  $\text{sc}(L(A_1)L(B_1)) = 2\beta = \alpha$ .

Now, let  $2m + 2n + 1 \leq \alpha \leq f(m, n+1) - 1$  and  $\alpha$  be odd. Let  $\beta = (\alpha+1)/2$ . Then  $m + n + 1 \leq \beta \leq f(m, n)$ , and we use the induction hypothesis and our Construction 2, to get automata  $A_2$  and  $B_2$  over  $\Sigma \cup \{a_n, b_n\}$  satisfying (1)–(4) such that  $\text{sc}(L(A_2)L(B_2)) = 2\beta - 1 = \alpha$ .

Finally, if  $m + (n+1) + 1 \leq \alpha \leq 2m + 2n$ , we set  $\beta = \alpha - 1$ . Then  $m+n+1 \leq \beta \leq f(m, n)$  since we have  $2m + 2n - 1 \leq m2^n - 2^{n-1}$  if  $m \geq 3$  and  $n \geq 2$ . We use the induction hypothesis and Construction 3, get appropriate automata  $A_3$  and  $B_3$ , satisfying (1)–(4) such that  $\text{sc}(L(A_3)L(B_3)) = \beta + 1 = \alpha$ . Our proof is complete.  $\square$

Now we consider the case of  $m = 2$  and  $n \geq 2$ . In such a case, we only need to modify conditions (1)–(4). All the proofs are the same as above, except for the base case, which is a bit more complicated in this case.

**Lemma 9.** *Let  $m = 2$ ,  $n \geq 2$ . For each  $\alpha$  with  $n+3 \leq \alpha \leq f(2, n) = 2^n + 2^{n-1}$ , there exist a minimal 2-state DFA  $A$ , and a minimal  $n$ -state DFA  $B$ , both defined over an alphabet  $\Sigma$  with  $|\Sigma| \leq 2n + 4$ , and such that  $\text{sc}(L(A)L(B)) = \alpha$ .*

*Proof.* We modify conditions (1)–(4) as follows.

- (1') The transitions on symbols  $a, b, c$  in states in  $\{q_0, q_1\} \cup \{0, 1\}$  are defined as in Fig. 1 for  $m = 2$ . This means that the subsets  $\{q_1, 0\}$ ,  $\{q_1, 0, 1\}$ , and  $\{q_0, 0, 1\}$  are reachable in  $D$ , that is, they are in  $\mathcal{R}$ .
- (2') If  $(q_i, a, q_0)$  is a transition in  $A$ , then  $i = 1$
- (3') Each set in  $\mathcal{R} \setminus \{\{q_0\}\}$  is reachable from  $\{q_1, 0\}$  in the subset automaton  $D$ .
- (4') For each state  $q$  of NFA  $N$ , there exists a string  $w_q$  in  $\Sigma^*$  which is accepted by  $N$  only from state  $q$ . Moreover, we have

$$\begin{aligned} w_1 &= \varepsilon, \\ w_0 &= c, \\ w_{q_1} &= bc, \\ w_{q_0} &= cbc, \\ w_j &= a_j \text{ for } j = 2, 3, \dots, n-1. \end{aligned}$$

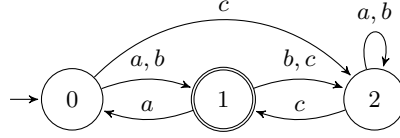
Now we continue with exactly the same constructions as in the case of  $m \geq 3$ , and, using induction on  $n$ , we get the lemma.  $\square$

The case of  $m = 1$  and  $n \geq 3$  is slightly different, although, the main idea is the same.

**Lemma 10.** *Let  $m = 1$  and  $n \geq 3$ . For each  $\alpha$  with  $n+1 \leq \alpha \leq f(1, n) = 2^{n-1}$ , there exist a minimal 1-state DFA  $A$ , and a minimal  $n$ -state DFA  $B$ , both defined over an alphabet  $\Sigma$  with  $|\Sigma| = n - 1$ , and such that  $\text{sc}(L(A)L(B)) = \alpha$ .*

*Proof (Proof Idea).* Let  $A$  be a 1-state DFA accepting  $\Sigma^*$ . We prove the lemma again by induction on  $n$ , where we assume that the following conditions hold for DFA  $B$ , the NFA  $N$  for  $\Sigma^*B$ , constructed from  $B$  by adding a loop in the initial state 0 on each input symbol in  $\Sigma$ , for the subset automaton  $D$  of  $N$ , and the set  $\mathcal{R}$  of reachable subsets in  $D$ :

- (1'') In DFA  $B$ , the transitions on  $a, b, c$  in states 0, 1, 2 are as in Fig. 2.
- (2'') In DFA  $B$ , we have  $0 \cdot \sigma \neq 0$  for each  $\sigma \in \Sigma$ .
- (3'') Each subset in  $\mathcal{R}$ , except for the initial subset  $\{0\}$ , can be reached from the subset  $\{0, 1\}$ .
- (4'') NFA  $N$  satisfies the condition in Proposition 1, that is, for each state  $j$  of  $N$ , there exists a string  $w_j$  in  $\Sigma^*$  which is accepted by  $N$  only from state  $j$ . Moreover, we have  $w_0 = c$  and  $w_1 = \varepsilon$ .



**Fig. 2.** Base case if  $m = 1$ .

The basis, in which we have  $n = 3$  and  $n + 1 = f(1, 3) = 4$ , holds true since the 3-state DFA  $B$  shown in Fig. 2 satisfies (1'')–(4'').

For the induction step, we again describe three constructions: We construct  $(n + 1)$ -state DFAs  $B_1, B_2, B_3$  from DFA  $B$  by adding a new state  $n$ , and by adding transitions on new symbol  $a_n, b_n$ , as shown in Table 2 in columns C1, C2, and C3, respectively.

We can show that all the resulting automata satisfy conditions (1'')–(4''), and, moreover, if  $|\mathcal{R}| = \beta$ , then  $|\mathcal{R}_1| = 2\beta$ ,  $|\mathcal{R}_2| = 2\beta - 1$ ,  $|\mathcal{R}_3| = \beta + 1$ . Since  $N$  and  $N_i$  satisfy (4''), we have  $\text{sc}(L(A)L(B)) = |\mathcal{R}|$  and  $\text{sc}(L(A_i)L(B_i)) = |\mathcal{R}_i|$ . This proves the lemma by induction.  $\square$

Up to now we have considered the complexities in the range from  $m + n + 1$  to  $f(m, n)$ . The complexities from 1 to  $m + n - 1$  are covered by the following result from [8]. Notice that this lemma also covers the case of  $m = 1$  and  $n = 2$ , since then  $f(1, 2) = 2^1 = 2 = m + n - 1$ .

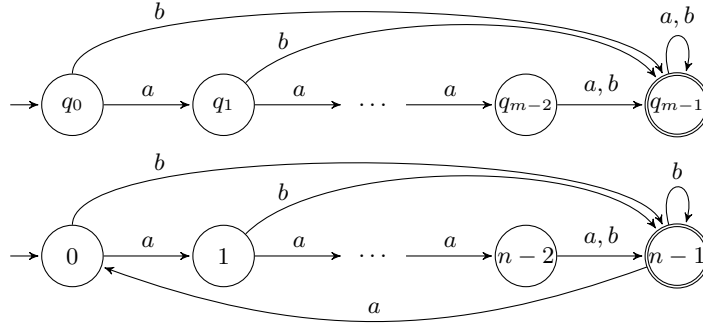
**Lemma 11 ([8, Lemma 5]).** *Let  $m, n \geq 1$ . For each  $\alpha$  with  $1 \leq \alpha \leq m + n - 1$ , there exist a minimal  $m$ -state DFA  $A$  and a minimal  $n$ -state DFA  $B$ , both defined over an alphabet of at most two symbols, such that  $\text{sc}(L(A)L(B)) = \alpha$ .*

The next lemma shows that the complexity  $m + n$  can be produced. Then we consider the case of  $n = 1$ .

**Lemma 12.** *Let  $m \geq 2, n \geq 2$ . There exist binary regular languages  $K$  and  $L$  with  $\text{sc}(K) = m$  and  $\text{sc}(L) = n$  such that  $\text{sc}(KL) = m + n$ .*

	C1	C2	C3
$\sigma \in \Sigma$	$n \rightarrow n$	$n \rightarrow n$	$n \rightarrow 0$
$a_n$	$n \rightarrow 1$	$n \rightarrow 2$	$n \rightarrow 1$
	$j \rightarrow n$	$0 \rightarrow 1$	$j \rightarrow 2$
		$1 \rightarrow n$	
		$j \rightarrow 0$ if $j \geq 2$	
$b_n$	–	–	$n \rightarrow n$
			$j \rightarrow n$
$w_n$	$a_n$	$a_n c$	$a_n$

**Table 2.** The three constructions in the case of  $m = 1$ .



**Fig. 3.** The minimal DFAs  $A$  and  $B$  with  $\text{sc}(L(A)L(B)) = m + n$ .

*Proof.* Let  $K$  and  $L$  be the binary languages accepted by minimal DFAs  $A$  and  $B$  shown in Fig. 3, where for each  $i$  in  $\{0, 1, \dots, m - 1\}$  and  $j$  in  $\{0, 1, \dots, n - 1\}$ , we have

$$q_i \cdot a = q_{i+1} \text{ if } i \neq m - 1, q_{m-1} \cdot a = q_{m-1} \text{ and } q_i \cdot b = q_{m-1};$$

$$j \cdot a = j + 1 \text{ if } j \neq n - 1, (n - 1) \cdot a = 0, \text{ and } j \cdot b = n - 1.$$

Construct an NFA  $N$  from DFAs  $A$  and  $B$  by adding transitions  $(q_{m-2}, a, 0)$ ,  $(q_{m-1}, a, 0)$ , and  $(q_i, b, 0)$  for each  $i$ ; the initial state of  $N$  is  $q_0$ , and the set of final states is  $\{n - 1\}$ . In the corresponding subset automaton, the initial subset is  $\{q_0\}$ , and we have

$$\{q_0\} \xrightarrow{a^i} \{q_i\} \text{ for } i = 1, 2, \dots, m - 2,$$

$$\{q_{m-2}\} \xrightarrow{a} \{q_{m-1}, 0\} \xrightarrow{a^j} \{q_{m-1}, 0, 1, \dots, j\} \text{ for } j = 1, 2, \dots, n - 1, \text{ and}$$

$$\{q_{m-1}, 0\} \xrightarrow{b} \{q_{m-1}, 0, n - 1\}.$$

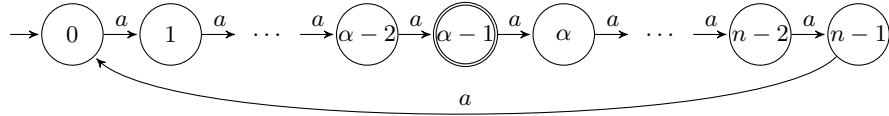
It follows that the subset automaton has  $m + n$  reachable subsets. Notice that each of these  $m + n$  subsets goes to some of them by  $a$ , and each of them goes to  $\{q_{m-1}, 0\}$  or to  $\{q_{m-1}, 0, n - 1\}$  by  $b$ . By Proposition 3, no other set is reachable.

To prove distinguishability, let  $\{q_i\} \cup S$  and  $\{q_j\} \cup T$  be two distinct reachable subsets. Since NFA  $N$  accepts the string  $a^{n-1-t}$  only from state  $t$  ( $0 \leq t \leq n - 1$ ), the two subsets are distinguishable if  $S \neq T$ . Next, if  $i < j$ , then we have

$$\{q_j\} \xrightarrow{a^{m-1-j}} \{q_{m-1}, 0\} \text{ and } \{q_i\} \xrightarrow{a^{m-1-j}} \{q_{m-1-(j-i)}\},$$

where the resulting subsets are distinguishable since they differ in a state of  $B$ . This proves distinguishability and concludes the proof.  $\square$

**Lemma 13.** *Let  $m \geq 1$  and  $n = 1$ . For each integer  $\alpha$  with  $1 \leq \alpha \leq f(m, 1) = m$ , there exist a minimal  $m$ -state DFA  $A$  and a minimal 1-state DFA  $B$ , both defined over a unary alphabet, such that  $\text{sc}(L(A)L(B)) = \alpha$ .*



**Fig. 4.** The minimal DFA  $A$  with  $\text{sc}(L(A)\Sigma^*) = \alpha$ .

*Proof.* Let  $A$  be a minimal  $m$ -state DFA shown in Fig. 4 accepting the language  $a^{\alpha-1}(a^m)^*$ . Let  $B$  be the minimal 1-state DFA accepting the unary language  $a^*$ . Then  $L(A)L(B) = a^{\alpha-1}(a^m)^*a^* = \{a^k \mid k \leq \alpha - 1\}$ , so  $\text{sc}(L(A)L(B)) = \alpha$ .  $\square$

The next theorem summarizes our results, and shows that the whole range of complexities for the concatenation operation can be produced using an alphabet which grows linearly with  $n$ . Recall that  $f(m, n)$  is the state complexity of the concatenation operation on languages over an alphabet of size at least two and we have  $f(m, 1) = m$  and  $f(m, n) = (m - 1)2^n + 2^{n-1}$  if  $n \geq 2$ .

**Theorem 14.** *Let  $m, n \geq 1$ . For each  $\alpha$  with  $1 \leq \alpha \leq f(m, n)$ , there exist regular languages  $K$  and  $L$  defined over an alphabet  $\Sigma$  with  $|\Sigma| \leq 2n + 4$  such that  $\text{sc}(K) = m$ ,  $\text{sc}(L) = n$ , and  $\text{sc}(KL) = \alpha$ .*

*Proof.* In each of the following six cases, we refer to the corresponding lemma dealing with this case:

- (1) If  $n = 1$ , then  $f(m, n) = m$ , and the theorem follows by Lemma 13.
- (2) If  $n \geq 2$  and  $1 \leq \alpha \leq m + n - 1$ , then the theorem follows by Lemma 11.
- (3) If  $m = 1$  and  $n = 2$ , then  $f(1, 2) = 2 = m + n - 1$ , so this case is covered by Lemma 11 as well.
- (4) If  $m = 1$ ,  $n \geq 3$ , and  $m + n = n + 1 \leq \alpha \leq f(1, n) = 2^{n-1}$ , then the theorem follows by Lemma 10.
- (5) The case of  $m \geq 2$ ,  $n \geq 2$ , and  $\alpha = m + n$  follows by Lemma 12.
- (6) Finally, if  $m \geq 2$ ,  $n \geq 2$ , and  $m + n + 1 \leq \alpha \leq f(m, n)$ , then the theorem follows by Lemma 9 if  $m = 2$ , and by Lemma 8 if  $m \geq 3$ .

This covers all the possible cases, and proves the theorem.  $\square$

## 5 Conclusions

We investigated the state complexity of languages resulting from the concatenation operation. We proved that for all  $m, n, \alpha$  with  $m, n \geq 1$  and  $1 \leq \alpha \leq f(m, n)$ , where  $f(m, n)$  is the state complexity of the concatenation operation, there exist regular languages  $K$  and  $L$  defined over an alphabet of size at most  $2n + 4$  such that  $\text{sc}(K) = m$ ,  $\text{sc}(L) = n$ , and  $\text{sc}(KL) = \alpha$ . This improves the result from [8], where an alphabet of size growing exponentially with  $n$  is used to produce the whole range of complexities for the concatenation operation. Our result complements similar results from [10, 14], where a linear alphabet is used to get the whole range of complexities for the reversal and Kleene closure operations.

A similar problem for the square operation, defined as  $L^2 = LL$ , remains open even for an exponential alphabet.

## References

1. Geffert, V.: Magic numbers in the state hierarchy of finite automata. *Inform. Comput.* 205, 1652–1670 (2007)
2. Holzer, M., Jakobi, S., Kutrib, M.: The magic number problem for subregular language families. *Internat. J. Found. Comput. Sci.* 23, 115–131 (2012)
3. Hopcroft, J. E., Ullman, J. D.: *Introduction to automata theory and computation.* Addison-Wesley Publishing Company (1979)
4. Iwama, K., Kambayashi, Y., Takaki, K.: Tight bounds on the number of states of DFAs that are equivalent to  $n$ -state NFAs. *Theoret. Comput. Sci.* 237, 485–494 (2000)
5. Iwama, K., Matsuura, A., Paterson, M.: A family of NFAs which need  $2^n - \alpha$  deterministic states. *Theoret. Comput. Sci.* 301, 451–462 (2003)
6. Jirásek, J., Jirásková, G., Szabari, A.: State complexity of concatenation and complementation. *Internat. J. Found. Comput. Sci.* 16, 511–529 (2005)
7. Jirásková, G.: The ranges of state complexities for complement, star, and reversal of regular languages. *Internat. J. Found. Comput. Sci.* 25, 101–124 (2014). Preliminary version: On the state complexity of complements, stars, and reversals of regular languages. In: Ito M., Toyama M. (eds.) *DLT 2008. LNCS*, vol. 5257, pp. 431–442. Springer (2008)
8. Jirásková, G.: Concatenation of regular languages and descriptive complexity. *Theory Comput. Syst.* 49, 306–318 (2011)
9. Jirásková, G.: Magic numbers and ternary alphabet. *Internat. J. Found. Comput. Sci.* 22, 331–344 (2011)
10. Jirásková, G., Palmovský, M., Šebej, J.: Kleene closure on regular and prefix-free languages. In: Holzer, M., Kutrib, M. (eds.) *CIAA 2014. LNCS*, vol. 8587, pp. 226–237. Springer (2014)
11. Maslov, A. N.: Estimates of the number of states of finite automata. *Soviet Math. Doklady* 11, 1373–1375 (1970)
12. Rabin, M., Scott, D.: Finite automata and their decision problems. *IBM Res. Develop.* 3, 114–129 (1959)
13. Sipser, M.: *Introduction to the theory of computation.* PWS Publishing Company, Boston (1997)
14. Šebej, J.: Reversal on regular languages and descriptive complexity. In: Jürgensen, H., Reis, R. (eds.) *DCFS 2013. LNCS*, vol. 8031, pp. 265–276. Springer (2013)
15. Yu, S.: Chapter 2: Regular languages. In: Rozenberg, G., Salomaa, A. (eds.) *Handbook of Formal Languages - Vol. I*, pp. 41–110. Springer (1997)
16. Yu, S., Zhuang, Q., Salomaa, K.: The state complexity of some basic operations on regular languages. *Theoret. Comput. Sci.* 125, 315–328 (1994)
17. Zijl, L.: Magic numbers for symmetric difference NFAs. *Internat. J. Found. Comput. Sci.* 16, 1027 (2005)
18. Čevorová, K.: Kleene star on unary regular languages. In: Jürgensen, H., Reis, R. (eds.) *DCFS 2013. LNCS*, vol. 8031, pp. 277–288. Springer (2013)
19. Čevorová, K., Jirásková, G., Krajňáková, I.: On the square of regular languages. In: Holzer, M., Kutrib, M. (eds.) *CIAA 2014. LNCS*, vol. 8587, pp. 136–147. Springer (2014)