

# Nondeterministic Complexity of Operations on Free and Convex Languages

Michal Hospodár<sup>(✉)</sup>, Galina Jirásková, and Peter Mlynárčik

Mathematical Institute, Slovak Academy of Sciences,  
Grešákova 6, 040 01 Košice, Slovakia  
hosmich@gmail.com, jiraskov@saske.sk, mlynarcik1972@gmail.com

**Abstract.** We study the nondeterministic state complexity of basic regular operations on the classes of prefix-, suffix-, factor-, and subword-free and -convex regular languages. For the operations of intersection, union, concatenation, square, star, reversal, and complementation, we get the tight upper bounds for all considered classes except for complementation on factor- and subword-convex languages. Most of our witnesses are described over optimal alphabets. The most interesting result is the describing of a proper suffix-convex language over a five-letter alphabet meeting the upper bound  $2^n$  for complementation.

## 1 Introduction

The nondeterministic state complexity of a regular language  $L$ ,  $\text{nsc}(L)$ , is the smallest number of states in any nondeterministic finite automaton (NFA) with a single initial state accepting the language  $L$ . The nondeterministic state complexity of a regular operation is defined as the maximal nondeterministic state complexity of languages resulting from the operation, considered as a function of nondeterministic state complexities of the operands. The languages that meet this maximal complexity for an operation are called witnesses for the operation. The (deterministic) state complexity of a regular language and a regular operation are defined analogously.

If operands for an operation satisfy some additional properties, the resulting complexity may be smaller than in the general case. In this paper we focus on the classes of prefix-, suffix-, factor-, subword-free and -convex languages. In the deterministic case, the complexity of basic regular operations on the classes of closed, ideal, and free languages was examined by Brzozowski *et al.* [2, 4, 5]. Some partial results in the classes of convex languages can be found in [3].

The nondeterministic state complexity of basic operations on regular languages was investigated by Holzer and Kutrib [9], and binary witnesses for reversal and complementation were presented in [11]. Han *et al.* [7, 8] studied

---

Research supported by VEGA grant 2/0084/15 and grant APVV-15-0091. This work was conducted as a part of PhD study of Michal Hospodár and Peter Mlynárčik at the Faculty of Mathematics, Physics and Informatics of the Comenius University.

the nondeterministic complexity of operations on prefix-free and suffix-free languages; some of their results were improved in [13]. Mlynárčik [15] examined the nondeterministic complexity of complementation on the classes of free and ideal languages. The remaining operations on ideal languages as well as all basic operations on closed languages were investigated in [10].

In this paper we continue this research and study the nondeterministic complexity of operations of intersection, union, concatenation, square, star, reversal, and complementation on the classes of prefix-, suffix-, factor-, subword-free and -convex languages. Except for complementation on factor- and subword-convex languages, we provide tight upper bounds, and to prove tightness, we use a small fixed alphabet in most cases. In some cases, we improve the known results by decreasing the size of alphabet for witness languages. We fix a small bug from the literature [8, Theorem 3.2] concerning union on prefix-free languages.

## 2 Preliminaries

We use a standard model of a nondeterministic finite automaton (NFA), as explained, for example, in [16]. An NFA  $A = (Q, \Sigma, \cdot, s, F)$  is a (complete) *deterministic finite automaton* (DFA) if for each state  $q$  in  $Q$  and each input symbol  $a$  in  $\Sigma$ , the set  $q \cdot a$  has exactly one element. If  $|q \cdot a| \leq 1$  for each  $q$  and  $a$ , then  $A$  is an *incomplete* DFA. In an  $\varepsilon$ -NFA, we also allow the transitions on the empty string. It is known that for each  $\varepsilon$ -NFA there exists an equivalent NFA with the same number of states [17, Theorem 2.3]. Sometimes, we allow an NFA to have multiple initial states and use the shortcut NNFA for this model.

A state  $q$  of an NFA  $A$  is called a *dead state* if no string is accepted by  $A$  from  $q$ , that is, if  $q \cdot w \cap F = \emptyset$  for each string  $w$ . An NFA  $A$  is a *trim* NFA if each its state  $q$  is reachable, that is, there is a string  $u$  in  $\Sigma^*$  such that  $q \in s \cdot u$ , and, moreover, no state of  $A$  is dead. We say that  $(p, a, q)$  is a transition in NFA  $A$  if  $q \in p \cdot a$ . We also say that the state  $q$  has an *in-transition* on symbol  $a$ , and the state  $p$  has an *out-transition* on symbol  $a$ . An NFA is *non-returning* if its initial state does not have any in-transitions, and it is *non-exiting* if each final state of  $A$  does not have any out-transitions.

**Definition 1.** A set of pairs of strings  $\{(u_1, v_1), (u_2, v_2), \dots, (u_n, v_n)\}$  is called a fooling set for a language  $L$  if for all  $i, j$  in  $\{1, 2, \dots, n\}$ ,

- (F1)  $u_i v_i \in L$ , and
- (F2) if  $i \neq j$ , then  $u_i v_j \notin L$  or  $u_j v_i \notin L$ .

**Lemma 2** ([1, Lemma 1]). Let  $\mathcal{F}$  be a fooling set for a language  $L$ . Then every NNFA for the language  $L$  has at least  $|\mathcal{F}|$  states. □

Let us emphasize that the size of a fooling set for  $L$  provides a lower bound on the number of states in any NNFA for  $L$ . If we insist on having just one initial state, then the following modification of a fooling set method can be used.

**Lemma 3** ([12, Lemma 4]). *Let  $\mathcal{A}$  and  $\mathcal{B}$  be sets of pairs of strings and let  $u$  and  $v$  be two strings such that  $\mathcal{A} \cup \mathcal{B}$ ,  $\mathcal{A} \cup \{(\varepsilon, u)\}$ , and  $\mathcal{B} \cup \{(\varepsilon, v)\}$  are fooling sets for a language  $L$ . Then every NFA for  $L$  has at least  $|\mathcal{A}| + |\mathcal{B}| + 1$  states.  $\square$*

Let  $A = (Q, \Sigma, \cdot, I, F)$  be an NNFA and  $S, T \subseteq Q$ . We say that  $S$  is *reachable* in  $A$  if there is a string  $w$  in  $\Sigma^*$  such that  $S = I \cdot w$ . Next, we say that  $T$  is *co-reachable* in  $A$  if  $T$  is reachable in  $A^R$  obtained from  $A$  by swapping the role of initial and final states and by reversing all the transitions. The next two observations are used throughout this paper.

**Lemma 4.** *Let  $A$  be an NNFA. Let for each state  $q$  of  $A$ , the singleton set  $\{q\}$  be reachable as well as co-reachable in  $A$ . Then  $A$  is minimal.*

*Proof.* Let  $A = (Q, \Sigma, \cdot, I, F)$ . Since  $\{q\}$  is reachable in  $A$ , there is a string  $u_q$  such that  $I \cdot u_q = \{q\}$ . Since  $\{q\}$  is co-reachable in  $A$ , there is a string  $v_q$  accepted by  $A$  from and only from the state  $q$ . Then  $\{(u_q, v_q) \mid q \in Q\}$  is a fooling set for  $L(A)$ . By Lemma 2, the NNFA  $A$  is minimal.  $\square$

**Lemma 5.** *Let  $A$  be a trim NFA. If both  $A$  and  $A^R$  are incomplete DFAs, then  $A$  and  $A^R$  are minimal NFAs.*

*Proof.* If  $A$  is a trim incomplete DFA, then for each state  $q$  of  $A$ , the singleton set  $\{q\}$  is reachable. If moreover  $A^R$  is an incomplete DFA, then  $\{q\}$  is co-reachable in  $A$ . By Lemma 4,  $A$  and  $A^R$  are minimal NFAs.  $\square$

If  $u, v, w, x \in \Sigma^*$  and  $w = uxv$ , then  $u$  is a *prefix* of  $w$ ,  $x$  is a *factor* of  $w$ , and  $v$  is a *suffix* of  $w$ . If  $w = u_0v_1u_1 \cdots v_nu_n$ , where  $u_i, v_i \in \Sigma^*$ , then  $v_1v_2 \cdots v_n$  is a *subword* of  $w$ . A prefix  $v$  (suffix, factor, subword) of  $w$  is *proper* if  $v \neq w$ .

A language  $L$  is *prefix-free* if  $w \in L$  implies that no proper prefix of  $w$  is in  $L$ ; it is *prefix-closed* if  $w \in L$  implies that each prefix of  $w$  is in  $L$ ; and it is *prefix-convex* if  $u, w \in L$  and  $u$  is a prefix of  $w$  imply that each string  $v$  such that  $u$  is a prefix of  $v$  and  $v$  is a prefix of  $w$  is in  $L$ . Suffix-, factor-, and subword-free, -closed, and -convex languages are defined analogously. It is known that a minimal NFA for a prefix-free (suffix-free) language is non-exiting (non-returning) [7, 8]. The next lemma gives a sufficient condition for an incomplete DFA accepting a suffix-free language.

**Lemma 6.** ([6, Lemma 1]). *Let  $A$  be a non-returning incomplete DFA that has a unique final state. If each state of  $A$  has at most one in-transition on every input symbol, then  $L(A)$  is suffix-free.  $\square$*

### 3 Unary Convex Languages

We start with examination of unary free and unary convex languages. Notice that if  $i \leq j$ , then  $a^i$  is a prefix, suffix, factor, and subword of  $a^j$ . It follows that in the unary case, all free classes and all convex classes coincide. Moreover, if  $n \geq 2$  then  $L = \{a^{n-1}\}$  is the only unary free language with  $\text{nsc}(L) = n$ .

Let  $L$  be a unary convex language and  $k$  be the length of the shortest string in  $L$ . If  $L$  is infinite, then  $L = \{a^i \mid i \geq k\}$ . If  $L$  is finite and  $\ell$  is the length of the longest string in  $L$ , then  $L = \{a^i \mid k \leq i \leq \ell\}$ . The set  $\{(a^i, a^{t-i}) \mid 0 \leq i \leq t\}$  is a fooling set for  $L$ , where  $t = k$  for infinite  $L$  and  $t = \ell$  for finite  $L$ . Thus the minimal incomplete DFA for  $L$ , which has  $t + 1$  states, is a minimal NFA for  $L$ .

The next theorem provides tight upper bounds for unary convex languages. All the results, except for the intersection and complementation, hold true for free languages too. On unary free languages, the complexity of intersection is  $n$  if  $m = n$  and 1 if  $m \neq n$ , and the complexity of complementation is  $\Theta(\sqrt{n})$  [13].

**Theorem 7 (Unary Convex Languages).** *Let  $m, n \geq 2$ . Let  $K$  and  $L$  be unary convex languages with  $\text{nsc}(K) = m$  and  $\text{nsc}(L) = n$ . Then*

- (a)  $\text{nsc}(K \cap L), \text{nsc}(K \cup L) \leq \max\{m, n\}$ ,
- (b)  $\text{nsc}(KL) \leq m + n - 1$  and  $\text{nsc}(L^2) \leq 2n - 1$ ,
- (c)  $\text{nsc}(L^*) \leq n - 1$ ,  $\text{nsc}(L^R) \leq n$ , and  $\text{nsc}(L^c) \leq n + 1$ ,

and all these upper bounds are tight.

*Proof.* The upper bound for intersection and union can be verified by the case analysis, where  $K$  and  $L$  can be finite or infinite. The upper bounds for concatenation, square, and complementation follow from the fact that the minimal NFAs can be incomplete deterministic. The upper bound for reversal follows from the fact that  $L^R = L$ .

Now we prove an upper bound for star. Let  $L$  be a unary convex language with  $\text{nsc}(L) = n$ . If  $L$  is infinite, then  $L = a^{n-1}a^*$ , and the language  $L^*$  is accepted by the  $(n - 1)$ -state NFA  $N = (\{0, 1, \dots, n - 2\}, \{a\}, \cdot, 0, \{0\})$  where  $i \cdot a = \{i + 1\}$  if  $i < n - 2$  and  $i \cdot a = \{0, n - 2\}$  if  $i = n - 2$ .

If  $L$  is finite, then there is an integer  $k$  such that  $L = \{a^i \mid k \leq i \leq n - 1\}$ . Then the  $(n - 1)$ -state NFA for  $L^*$  can be constructed from a minimal incomplete DFA  $(\{0, 1, \dots, n - 1\}, \{a\}, \cdot, 0, \{k, k + 1, \dots, n - 1\})$  for  $L$  by making the state  $n - 1$  initial, adding the transition  $(n - 1, a, 1)$ , and removing the state 0.

The languages  $a^{m-1}a^*$  and  $a^{n-1}a^*$  meet the upper bound for intersection, the languages  $a^{m-1}$  and  $a^{n-1}$  meet the upper bound for union and concatenation, the language  $a^{n-1}$  meets the upper bound for square, star, and reversal, and the language  $\{a^i \mid i \leq n - 1\}$  meets the upper bound for complementation.  $\square$

Table 1 summarizes our results on unary convex languages and compares them to the known results on unary regular and free languages [9, 13].

**Table 1.** The nondeterministic complexity of operations on unary classes; the results are from [13] (first row), Theorem 7 (second row), and [9] (third row).

	$K \cap L$	$K \cup L$	$KL$	$L^2$	$L^*$	$L^c$
Unary free	$n; m = n$	$\max\{m, n\}$	$m + n - 1$	$2n - 1$	$n - 1$	$\Theta(\sqrt{n})$
Unary convex	$\max\{m, n\}$	$\max\{m, n\}$	$m + n - 1$	$2n - 1$	$n - 1$	$n + 1$
Unary regular	$mn; \gcd(m, n) = 1$	$m + n + 1;$ $\gcd(m, n) = 1$	$\geq m + n - 1$ $\leq m + n$	$\geq 2n - 1$ $\leq 2n$	$n + 1$	$2^{\Theta(\sqrt{n} \log n)}$

## 4 Prefix-, Suffix-, Factor-, Subword-Free Languages

The nondeterministic complexity of operations on prefix- and suffix-free languages was studied by Han *et al.* [7,8], where tight upper bounds were obtained for basic regular operations. Some of their results were improved by decreasing the size of the alphabet for witness languages in [14]. The aim of this section is to get tight upper bounds on the nondeterministic state complexity of basic regular operations on factor- and subword-free languages as they are shown in Table 2. We also fix a small bug in [8] concerning union on prefix-free languages. For tightness, we use a unary or binary alphabet in all cases except for intersection on subword-free languages, and these alphabets are always optimal. The size of alphabet is shown in the right part of each column. The dot denotes the same complexity as in the previous column. The results for complementation are from [13], and the ternary alphabet is also optimal here. We start with intersection.

**Theorem 8 (Intersection).** *Let  $K$  and  $L$  be regular languages over an alphabet  $\Sigma$  such that  $\text{nsc}(K) = m$  and  $\text{nsc}(L) = n$ .*

- (a) *If  $K$  and  $L$  are prefix-free (suffix-free) then  $\text{nsc}(K \cap L) \leq mn - (m + n - 2)$ , and the bound is tight if  $m \geq 4$ ,  $n \geq 2$ , and  $|\Sigma| \geq 2$ .*
- (b) *If  $K$  and  $L$  are factor-free, then  $\text{nsc}(K \cap L) \leq mn - 2(m + n - 3)$ , and the bound is tight if  $m \geq 5$ ,  $n \geq 3$ , and  $|\Sigma| \geq 2$ .*
- (c) *If  $m, n \geq 3$ , then there exist subword-free regular languages  $K$  and  $L$  over an  $(m + n - 5)$ -letter alphabet such that  $\text{nsc}(K) = m$ ,  $\text{nsc}(L) = n$ , and  $\text{nsc}(K \cap L) = mn - 2(m + n - 3)$ .*

*Proof.* We first prove the upper bounds. Let  $A$  and  $B$  be minimal NFAs for  $K$  and  $L$ , respectively. We may assume that the state sets of  $A$  and  $B$  are  $\{0, 1, \dots, m - 1\}$  and  $\{0, 1, \dots, n - 1\}$ , respectively, with the initial state 0 in both automata. Construct the product automaton  $A \times B$  for  $K \cap L$ .

If  $K$  and  $L$  are prefix-free with the final states  $m - 1$  and  $n - 1$  respectively, then all states in the last row and last column, except for  $(m - 1, n - 1)$ , are dead, so we can omit them. If  $K$  and  $L$  are suffix-free, then  $A$  and  $B$  are non-returning, so all states in the first row and first column, except for  $(0, 0)$ , are unreachable. Since every factor-free language is both prefix-free and suffix-free, all the three upper bounds follow from these observations.

**Table 2.** The nondeterministic state complexity of operations on free classes; complementation is from [13]. The dot means that the complexity is the same as in the previous column.

	Regular	$ \Sigma $	Prefix-free	$ \Sigma $	Suffix-free	$ \Sigma $	Factor-free	$ \Sigma $	Subword-free	$ \Sigma $
$K \cap L$	$mn$	2	$mn - (m+n-2)$	2	.	2	$mn - 2(m+n-3)$	2	.	$m+n-5$
$K \cup L$	$m+n+1$	2	$m+n$	2	$m+n-1$	2	$m+n-2$	2	.	2
$KL$	$m+n$	2	$m+n-1$	1	.	1	.	1	.	1
$L^2$	$2n$	2	$2n-1$	1	.	1	.	1	.	1
$L^*$	$n+1$	1	$n$	2	.	2	$n-1$	1	.	1
$L^R$	$n+1$	2	$n$	1	$n+1$	2	$n$	1	.	1
$L^c$	$2^n$	2	$2^{n-1}$	3	.	3	$2^{n-2} + 1$	3	.	$2^{n-2}$

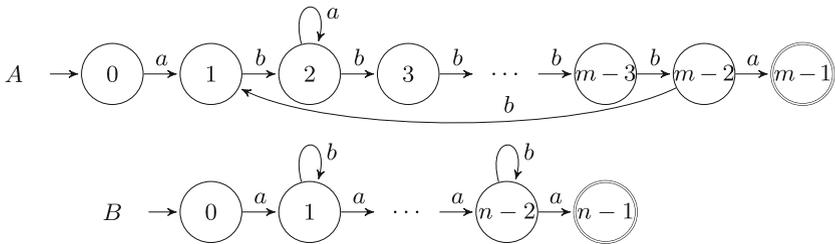
To prove tightness, we first consider factor-free languages. Let  $m \geq 5, n \geq 3$ . Let  $K$  and  $L$  be the languages accepted by the NFAs  $A$  and  $B$  shown in Fig. 1.

Every string  $w$  in  $K$  begins and ends with  $a$ , and  $|w|_b \bmod (m-2) = (m-3)$ . Every proper factor  $v$  of  $w$  which begins and ends with  $a$  has a computation in  $A$  which either starts in the state 0 and ends in the state 2, or starts and ends in 2, or starts in 2 and ends in  $m-1$ . However, in all three cases,  $|v|_b \bmod (m-2) \neq (m-3)$ , so  $v \notin L$ . Hence the language  $K$  is factor-free. Next, every string in  $L$  has exactly  $n-1$   $a$ 's, but every proper factor of every string in  $L$  has less than  $n-1$   $a$ 's. Hence  $L$  is factor-free.

Construct the product automaton  $A \times B$  and remove all the unreachable and dead states to get a trim NFA  $N$  for  $K \cap L$ . Since the NFA  $N$  and its reverse  $N^R$  are incomplete DFAs, the NFA  $N$  is minimal by Lemma 4. So we have  $\text{nsc}(K \cap L) = mn - 2(m+n-3)$ . Notice that there is no need to prove that NFAs  $A$  and  $B$  are minimal because the upper bound cannot be met by languages of a smaller complexity. For this reason we do not prove the minimality of witnesses in what follows.

Next, the left quotients of  $K$  and  $L$  by the string  $a$ , that is, the languages  $a \setminus K$  and  $a \setminus L$ , are prefix-free and meet the upper bound  $mn - (m+n-2)$ . Similarly, the right quotients  $K/a$  and  $L/a$  are suffix-free witnesses.

Finally, we consider the complexity of intersection on subword-free languages. Let  $\Sigma = \{a\} \cup \{b_k \mid 2 \leq k \leq m-2\} \cup \{c_\ell \mid 2 \leq \ell \leq n-2\}$ . Let  $K$  and  $L$  be



**Fig. 1.** Binary factor-free witnesses for intersection.

languages accepted by incomplete DFAs  $A = (\{0, 1, \dots, m - 1\}, \Sigma, \cdot, 0, \{m - 1\})$  and  $B = (\{0, 1, \dots, n - 1\}, \Sigma, \circ, 0, \{n - 1\})$ , where for each  $i$  ( $0 \leq i \leq m - 2$ ),  $j$  ( $0 \leq j \leq n - 2$ ),  $k$  ( $2 \leq k \leq m - 2$ ), and  $\ell$  ( $2 \leq \ell \leq n - 2$ ), we have

$$\begin{aligned}
 i \cdot a &= i + 1, & j \circ a &= j + 1, \\
 0 \cdot b_k &= k \text{ and } (k - 1) \cdot b_k = m - 1, & 0 \circ b_k &= 1 \text{ and } (n - 2) \circ b_k = n - 1, \\
 0 \cdot c_\ell &= 1 \text{ and } (m - 2) \cdot c_\ell = m - 1, & 0 \circ c_\ell &= \ell \text{ and } (\ell - 1) \circ c_\ell = n - 1.
 \end{aligned}$$

We can prove that  $K$  and  $L$  are subword-free and meet the upper bound for intersection.  $\square$

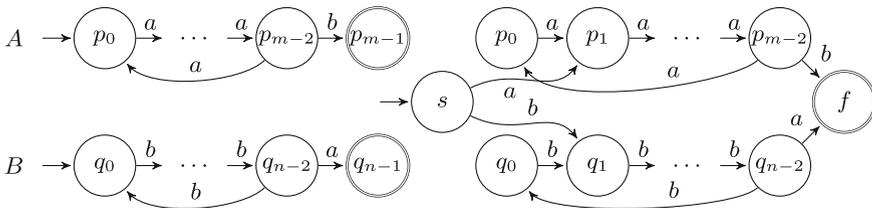
As a result of the previous theorem, we get the nondeterministic state complexity of intersection on each of the four classes of free languages, as it is shown in the corresponding row of Table 2.

Now we consider the union operation. In [8] it is claimed that the upper bound  $m + n$  is met by the union of prefix-free languages  $K = (a^{m-1})^*b$  and  $L = (c^{n-1})^*d$ , and a set  $P$  of pairs of strings of size  $m + n$  is described in [8, Proof of Theorem 3.2]. The authors claimed that  $P$  is a fooling set for  $K \cup L$ . However, the language  $K \cup L$  is accepted by an NNFA of  $m + n - 1$  states. Therefore  $P$  cannot be a fooling set for  $K \cup L$ . Here we prove the tightness of the upper bound  $m + n$  for union of prefix-free languages using a binary alphabet and a modified fooling set method given by Lemma 3. Next we get the tight upper bound for union of suffix-, factor-, and subword-free languages. To get tightness, we always use a binary alphabet which is optimal for all four classes.

**Theorem 9 (Union).** *Let  $K$  and  $L$  be regular languages over an alphabet  $\Sigma$  such that  $\text{nsc}(K) = m$  and  $\text{nsc}(L) = n$ .*

- (a) *If  $K$  and  $L$  are prefix-free then  $\text{nsc}(K \cup L) \leq m + n$ , and the bound is tight if  $m \geq 3, n \geq 3$ , and  $\Sigma \geq 2$ .*
- (b) *If  $K$  and  $L$  are suffix-free then  $\text{nsc}(K \cup L) \leq m + n - 1$ , and the bound is tight if  $m \geq 3, n \geq 3$ , and  $\Sigma \geq 2$ .*
- (c) *If  $K$  and  $L$  are factor-free, then  $\text{nsc}(K \cup L) \leq m + n - 2$ , and the bound is met by binary subword-free languages if  $m \geq 2$  and  $n \geq 2$ .*

*Proof.* We first prove the upper bounds. Let  $A$  and  $B$  be minimal NFAs for  $K$  and  $L$ , respectively, with disjoint state sets, and the initial states  $s_A$  and  $s_B$ , respectively.



**Fig. 2.** Binary prefix-free witnesses for union meeting the upper bound  $m + n$ .

- (a) If  $K$  and  $L$  are prefix-free, then NFAs  $A$  and  $B$  are non-exiting and have a unique final state. To get an  $(m + n)$ -state NFA for  $K \cup L$  from  $A$  and  $B$ , add a new initial (non-final) state connected through  $\varepsilon$ -transitions to  $s_A$  and  $s_B$ , make the states  $s_A$  and  $s_B$  non-initial, and merge the final states of  $A$  and  $B$ .
- (b) If  $K$  and  $L$  are suffix-free, then  $A$  and  $B$  are non-returning. We can get an  $(m + n - 1)$ -state NFA for  $K \cup L$  from  $A$  and  $B$  by merging their initial states.
- (c) If  $K$  and  $L$  are factor-free, then they are both prefix- and suffix-free. To get an  $(m + n - 2)$ -state NFA for  $K \cup L$  from  $A$  and  $B$ , we merge their initial states, and then we merge their final states.

To prove tightness, consider languages  $K$  and  $L$  accepted by an  $m$ -state and  $n$ -state NFAs  $A$  and  $B$ , respectively, shown in Fig. 2 (left). Notice that  $K$  is prefix-free since every string in  $K$  ends with  $b$  while every proper prefix of every string in  $K$  is in  $a^*$ . Similarly,  $L$  is prefix-free.

Construct the  $(m + n)$ -state NFA for their union by adding a new initial state  $s$ , by adding transitions  $(s, a, p_1)$  and  $(s, b, q_1)$ , by making states  $p_0$  and  $q_0$  non-initial, and by merging their final states as shown in Fig. 2 (right). The resulting trim NFA is an incomplete DFA, and its reverse is an incomplete DFA as well. By Lemma 5, this NFA is minimal. It follows that  $\text{nsc}(K \cup L) \geq m + n$ .

Next, the languages  $K^R$  and  $L^R$  are suffix-free, and they are accepted by  $m$ -state and  $n$ -state NFAs  $A^R$  and  $B^R$ , respectively. To get an NFA for  $K^R \cup L^R$ , we merge the initial states of  $A^R$  and  $B^R$ . For each state  $q$  of the resulting automaton, the singleton set  $\{q\}$  is reachable, as well as co-reachable. By Lemma 4, this NFA is minimal. Hence we get  $\text{nsc}(K^R \cup L^R) \geq m + n - 1$ . Finally, we again use Lemma 5 to show that the union of binary subword-free languages  $\{a^{m-1}\}$  and  $\{b^{n-1}\}$  meets the upper bound  $m + n - 2$ . □

The theorem above gives the nondeterministic state complexity of union on free languages, as it is shown in the corresponding row of Table 2.

The nondeterministic state complexity of concatenation on regular languages is  $m + n$  with binary witnesses [9, Theorem 7]. For prefix-free and suffix-free languages, the upper bound is  $m + n - 1$  [7, 8], and to prove tightness, a binary alphabet was used in [8, Theorem 3.1] and [7, Theorem 4]. In this section, we show that this upper bound is tight for all four classes of free languages, and to prove tightness, we use a unary alphabet.

**Theorem 10 (Concatenation, Square).** *Let  $K$  and  $L$  be prefix-free (suffix-free) languages with  $\text{nsc}(K) = m$  and  $\text{nsc}(L) = n$ . Then  $\text{nsc}(KL) \leq m + n - 1$ ,  $\text{nsc}(L^2) \leq 2n - 1$ , and these bounds are met by unary subword-free languages.*

*Proof.* Let  $A$  and  $B$  be minimal NFAs for  $K$  and  $L$ , respectively. In the prefix-free case, we can merge the final state of  $A$  and the initial state of  $B$  to get an NFA for  $KL$ . In the suffix-free case, automata  $A$  and  $B$  are non-returning. To get an NFA for  $KL$ , we add the transition  $(p, a, q)$  for each final state  $p$  of  $A$  and each transition  $(s_B, a, q)$  of  $B$ . Next, we make final states of  $A$  non-final,

and remove the unreachable state  $s_B$ . As a result, we get an NFA for  $KL$  of  $m + n - 1$  states in both cases. This upper bound is met by the concatenation of unary subword-free languages  $\{a^{m-1}\}$  and  $\{a^{n-1}\}$ . Since the witnesses have the same structure, the complexity  $2n - 1$  for square follows.  $\square$

Using the theorem above, we obtain the nondeterministic state complexity of concatenation and square on free languages, as shown in Table 2.

We next consider the Kleene star and reversal operations. Both operations have nondeterministic complexity  $n + 1$  on regular languages with a unary witness for star [9, Theorem 9] and a binary witness for reversal [11, Theorem 2]. The following theorem provides tight upper bounds for star on all four classes of free languages, as shown in Table 2. To get tightness, we use an optimal binary alphabet in the prefix- and suffix-free case, and a unary alphabet otherwise.

**Theorem 11 (Star).** *Let  $L$  be a language over an alphabet  $\Sigma$  with  $\text{nsc}(L) = n$ .*

- (a) *If  $L$  is prefix- or suffix-free then  $\text{nsc}(L^*) \leq n$ . These upper bounds are tight if  $|\Sigma| \geq 2$ , and the size of alphabet cannot be decreased.*
- (b) *If  $L$  is factor-free, then  $\text{nsc}(L^*) \leq n - 1$ , and the bound is met by a unary subword-free language.*

*Proof.* Let  $A = (Q, \Sigma, \cdot, s, F)$  be a minimal NFA for  $L$ .

- (a) If  $L$  is prefix-free, then  $A$  is non-exiting and has a unique final state  $q_f$ . We can construct an  $n$ -state  $\varepsilon$ -NFA for the language  $L^*$  from  $A$  by making state  $q_f$  initial and state  $s$  non-initial, and by adding the  $\varepsilon$ -transition from  $q_f$  to  $s$ . If  $L$  is suffix-free, then  $A$  is non-returning. Now we construct an  $n$ -state  $\varepsilon$ -NFA for  $L^*$  from  $A$  by making the initial state  $s$  final, and by adding the  $\varepsilon$ -transition from every final state to the initial state  $s$ . Suffix-free language  $a^{n-1}b^*$  and prefix-free language  $b^*a^{n-1}$  meet the upper bound  $n$ .
- (b) If  $L$  is factor-free, then  $A$  is non-returning, non-exiting, and it has a unique final state  $q_f$ . We construct an NFA for  $L^*$  by making state  $q_f$  initial, by adding transition  $(q_f, a, q)$  for each transition  $(s, a, q)$ , and by omitting the state  $s$ . The unary subword-free language  $\{a^{n-1}\}$  meets this upper bound.  $\square$

Now we turn our attention to the reversal operation. Han *et al.* obtained tight upper bounds for reversal on prefix-free and suffix-free languages and they provided a binary prefix-free witness [8, Theorem 3.4] and a ternary suffix-free witness [7, Theorem 9]. As shown in the next theorem, the upper bound for reversal on prefix-free languages is  $n$ , so it is met by any unary language. In the suffix-free case, we provide a binary witness meeting the upper bound  $n + 1$ . Notice that the reverse of a language accepted by an  $n$ -state NFA is accepted by an  $n$ -state NNFA. This means that we cannot use a fooling set method to prove the tightness of the bound  $n + 1$ . However, a modified fooling set method described in Lemma 3 can be successfully used here. As a result of this theorem, we get the corresponding row of Table 2.

**Theorem 12 (Reversal).**

- (a) Let  $L$  be a prefix-free language with  $\text{nsc}(L) = n$ . Then  $\text{nsc}(L^R) \leq n$ , and this bound is met by a unary subword-free language.
- (b) If  $n \geq 5$ , then there exists a binary suffix-free regular language  $L$  such that  $\text{nsc}(L) = n$  and  $\text{nsc}(L^R) = n + 1$ .

*Proof.*

- (a) If  $L$  is prefix-free, then every minimal NFA for  $L$  has a unique final state. Thus  $\text{nsc}(L^R) \leq n$ . The bound is met by the language  $\{a^{n-1}\}$ .
- (b) Set  $L = ba^{n-4}(a^{n-3})^* + ab(bb)^*$ . Since every string in  $L$  contain both  $a$  and  $b$ , but every proper suffix of every string in  $L$  is in  $a^* \cup b^*$ , the language  $L$  is suffix-free. Let
 
$$\mathcal{A} = \{(a^{n-3}, a^{n-4}b)\} \cup \{(a^i, a^{n-4-i}b) \mid 1 \leq i \leq n-4\} \cup \{(a^{n-4}b, \varepsilon)\},$$

$$\mathcal{B} = \{(bb, ba), (b, a)\},$$

$$u = ba, \text{ and } v = a^{n-4}b.$$

Using Lemma 3, we show that every NFA for  $L^R$  needs at least  $n + 1$  states.  $\square$

## 5 Convex Languages

In this section, we examine the nondeterministic state complexity of basic regular operations on convex languages. Recall that prefix-closed and right ideal languages are prefix-convex, and similar inclusions exist for suffix-, factor-, and subword-closed, and left, two-sided, and all-sided ideal classes.

The complexity of operations on closed and ideal languages was studied in [10], where for each operation, except for complementation, and for each of the four convex classes, a closed or an ideal witness, meeting the complexity of the operation in the class of regular languages, is described: binary subword-closed languages meeting the upper bound  $mn$  for intersection, and binary subword-closed witnesses meeting the upper bound  $m + n + 1$  for union are given in Theorems 4 and 5, ternary subword-closed languages meeting the bound  $m + n$  for concatenation and  $2n$  for square, are described in Theorem 6 and Corollary 7, the binary all-sided ideal meeting the upper bound  $n + 1$  for star is provided in Theorem 16, and for reversal, the binary prefix-closed, ternary factor-closed, and subword-closed witness defined over an alphabet of size  $2n - 2$  are described in Theorem 9. Therefore, as shown in Table 3, the complexity of operations on convex languages, except for complementation, is the same as for regular languages, although, to get tightness, larger alphabets are used in some cases.

The nondeterministic state complexity of complementation on regular languages is  $2^n$  with a binary witness [9, 11]. The upper bound  $2^n$  is met by binary prefix-closed languages [10, Theorem 10]. On the other hand, this upper bound cannot be met by suffix-closed, suffix-free, or left ideal languages [10, 13, 15].

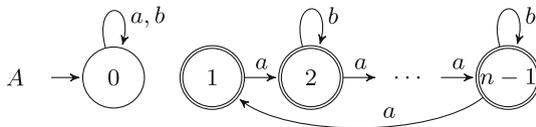
Our last result shows that the upper bound  $2^n$  for complementation is tight on suffix-convex languages. We describe a proper suffix-convex language, that is, a suffix-convex languages which is neither suffix-free, nor suffix-closed, nor left ideal, that meets this upper bound for complementation.

**Table 3.** The nondeterministic state complexity of operations on convex classes. The results for regular languages are from [9]. All the remaining results, except for complementation on suffix-convex languages, follow from [9, 10].

	Regular	$ \Sigma $	Prefix-convex	$ \Sigma $	Suffix-convex	$ \Sigma $	Factor-convex	$ \Sigma $	Subword-convex	$ \Sigma $
$K \cap L$	$mn$	2	.	2	.	2	.	2	.	2
$K \cup L$	$m + n + 1$	2	.	2	.	2	.	2	.	2
$KL$	$m + n$	2	.	3	.	3	.	3	.	3
$L^2$	$2n$	2	.	3	.	3	.	3	.	3
$L^*$	$n + 1$	1	.	2	.	2	.	2	.	2
$L^R$	$n + 1$	2	.	2	.	3	.	3	.	$2n - 2$
$L^c$	$2^n$	2	.	2	.	5	$\geq 2^{n-1} + 1 \leq 2^n$	2	$\geq 2^{n-1} + 1 \leq 2^n$	$2^n$

**Theorem 13 (Complementation on Suffix-Convex Languages).** *Let  $n \geq 3$ . There exists a suffix-convex regular language  $L$  over a 5-letter alphabet such that  $\text{nsc}(L) = n$  and  $\text{nsc}(L^c) = 2^n$ .*

*Proof.* Let  $L$  be the language accepted by the nondeterministic finite automaton  $A = (\{0, 1, \dots, n - 1\}, \{a, b, c, d, e\}, \cdot, 0, \{1, 2, \dots, n - 1\})$ , where the transitions on  $a$  and  $b$  are shown in Fig. 3, the transitions on  $c, d, e$  are as follows:  $0 \cdot c = \{0, 1, \dots, n - 1\}$ ,  $0 \cdot d = \{1, 2, \dots, n - 1\}$ ,  $q \cdot e = \{n - 1\}$  for each state  $q$  of  $A$ , and all the remaining transitions go to the empty set. In the NFA  $A^R$ , the final state 0 goes to itself on  $a, b, c$  and to the empty set on  $d$  and  $e$ . Next, every other state of  $A^R$  goes to 0 on  $d$ , and the state  $n - 1$  goes to  $\{0, 1, \dots, n - 1\}$  on  $e$ . Thus in the subset automaton of  $A^R$ , each final subset, that is, a subset containing the state 0, goes either to a final subset containing 0 or to the empty set on each input symbol. It follows that the language  $L^R$  is prefix-convex, so  $L$  is suffix-convex. We can show that each subset of the state set of  $A$  is reachable and co-reachable. Hence for each subset  $S$ , there exists a string  $u_S$  in  $\Sigma^*$  such that  $s \cdot u_S = S$ . Next,  $S^c$  is co-reachable, so there is a string  $v_S$  which is accepted by  $A$  from each state in  $S^c$ , but rejected from each state in  $S$ . Thus  $\{(u_S, v_S) \mid S \subseteq Q\}$  is a fooling set for  $L^c$  of size  $2^n$ , so  $\text{nsc}(L^c) \geq 2^n$  by Lemma 2.  $\square$



**Fig. 3.** Transitions on  $a$  and  $b$  in suffix-convex witness for complementation.

## 6 Conclusions

We investigated the nondeterministic state complexity of basic operations on the classes of prefix-, suffix-, factor-, subword-free and -convex languages. For each class and for each operation, except for complementation on factor- and subword-convex languages, we obtained the tight upper bounds.

Our results are summarized in Tables 1, 2, and 3. For complementation on factor- and subword-convex languages we do not know whether or not the upper bound  $2^n$  is tight. All the remaining upper bounds are tight. Whenever we used a binary alphabet, it is always optimal in the sense that the upper bound is not tight for any smaller alphabet. In any other case, we do not know whether the upper bounds are tight for a smaller alphabet. The complexity of complementation on factor- and subword-convex languages remains open as well.

**Acknowledgment.** We would like to thank Jozef Jirásek, Jr., for his help with finding the suffix-convex witness for complementation and for fruitful discussions on the topic.

## References

1. Birget, J.: Intersection and union of regular languages and state complexity. *Inform. Process. Lett.* **43**(4), 185–190 (1992). doi:[10.1016/0020-0190\(92\)90198-5](https://doi.org/10.1016/0020-0190(92)90198-5)
2. Brzozowski, J., Jirásková, G., Li, B., Smith, J.: Quotient complexity of bifix-, factor-, and subword-free regular languages. *Acta Cybernetica* **21**(4), 507–527 (2014)
3. Brzozowski, J.: Complexity in convex languages. In: Dediu, A.-H., Fernau, H., Martín-Vide, C. (eds.) *LATA 2010*. LNCS, vol. 6031, pp. 1–15. Springer, Heidelberg (2010). doi:[10.1007/978-3-642-13089-2\\_1](https://doi.org/10.1007/978-3-642-13089-2_1)
4. Brzozowski, J.A., Jirásková, G., Li, B.: Quotient complexity of ideal languages. *Theoret. Comput. Sci.* **470**, 36–52 (2013). doi:[10.1016/j.tcs.2012.10.055](https://doi.org/10.1016/j.tcs.2012.10.055)
5. Brzozowski, J.A., Jirásková, G., Zou, C.: Quotient complexity of closed languages. *Theory Comput. Syst.* **54**(2), 277–292 (2014). doi:[10.1007/s00224-013-9515-7](https://doi.org/10.1007/s00224-013-9515-7)
6. Cmorik, R., Jirásková, G.: Basic operations on binary suffix-free languages. In: Kotásek, Z., Bouda, J., Černá, I., Sekanina, L., Vojnar, T., Antoš, D. (eds.) *MEMICS 2011*. LNCS, vol. 7119, pp. 94–102. Springer, Heidelberg (2012). doi:[10.1007/978-3-642-25929-6\\_9](https://doi.org/10.1007/978-3-642-25929-6_9)
7. Han, Y., Salomaa, K.: Nondeterministic state complexity for suffix-free regular languages. In: McQuillan, I., Pighizzini, G. (eds.) *DCFS 2010*. EPTCS, vol. 31, pp. 189–196 (2010). doi:[10.4204/EPTCS.31.21](https://doi.org/10.4204/EPTCS.31.21)
8. Han, Y., Salomaa, K., Wood, D.: Nondeterministic state complexity of basic operations for prefix-free regular languages. *Fundam. Inform.* **90**(1–2), 93–106 (2009). doi:[10.3233/FI-2009-0008](https://doi.org/10.3233/FI-2009-0008)
9. Holzer, M., Kutrib, M.: Nondeterministic descriptonal complexity of regular languages. *Int. J. Found. Comput. Sci.* **14**(6), 1087–1102 (2003). doi:[10.1142/S0129054103002199](https://doi.org/10.1142/S0129054103002199)
10. Hospodár, M., Jirásková, G., Mlynářčík, P.: Nondeterministic complexity of operations on closed and ideal languages. In: Han, Y.-S., Salomaa, K. (eds.) *CIAA 2016*. LNCS, vol. 9705, pp. 125–137. Springer, Cham (2016). doi:[10.1007/978-3-319-40946-7\\_11](https://doi.org/10.1007/978-3-319-40946-7_11)

11. Jirásková, G.: State complexity of some operations on binary regular languages. *Theoret. Comput. Sci.* **330**(2), 287–298 (2005). doi:[10.1016/j.tcs.2004.04.011](https://doi.org/10.1016/j.tcs.2004.04.011)
12. Jirásková, G., Masopust, T.: Complexity in union-free regular languages. *Internat. J. Found. Comput. Sci.* **22**(7), 1639–1653 (2011). doi:[10.1142/S0129054111008933](https://doi.org/10.1142/S0129054111008933)
13. Jirásková, G., Mlynárčik, P.: Complement on prefix-free, suffix-free, and non-returning NFA languages. In: Jürgensen, H., Karhumäki, J., Okhotin, A. (eds.) *DCFS 2014*. LNCS, vol. 8614, pp. 222–233. Springer, Cham (2014). doi:[10.1007/978-3-319-09704-6\\_20](https://doi.org/10.1007/978-3-319-09704-6_20)
14. Jirásková, G., Olejár, P.: State complexity of intersection and union of suffix-free languages and descriptive complexity. In: Bordihn, H., et al. (eds.) *NCMA 2009*, vol. 256, pp. 151–166. Austrian Computer Society (2009). [books.ocg.at](http://books.ocg.at)
15. Mlynárčik, P.: Complement on free and ideal languages. In: Shallit, J., Okhotin, A. (eds.) *DCFS 2015*. LNCS, vol. 9118, pp. 185–196. Springer, Cham (2015). doi:[10.1007/978-3-319-19225-3\\_16](https://doi.org/10.1007/978-3-319-19225-3_16)
16. Sipser, M.: *Introduction to the Theory of Computation*. PWS Publishing Company, Boston (1997)
17. Yu, S.: Regular languages. In: Rozenberg, G., Salomaa, A. (eds.) *Handbook of Formal Languages*, vol. 1, pp. 41–110. Springer, Heidelberg (1997)