On the State Complexity of Complements, Stars, and Reversals of Regular Languages*

Galina Jirásková

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

Abstract. We examine the deterministic and nondeterministic state complexity of complements, stars, and reversals of regular languages. Our results are as follows:

1. The nondeterministic state complexity of the complement of an n-state NFA language over a five-letter alphabet may reach each value in the range from $\log n$ to 2^n.
2. The state complexity of the star (reversal) of an n-state DFA language over a growing alphabet may reach each value in the range from 1 to $\frac{3}{4}2^n$ (from $\log n$ to 2^n, respectively).
3. The nondeterministic state complexity of the star (reversal) of an n-state NFA binary language may reach each value in the range from 1 to $n+1$ (from $n-1$ to $n+1$, respectively).

We also obtain some partial results on the nondeterministic state complexity of the complements of binary regular languages. As a bonus, we get an exponential number of values that are non-magic, which improves a similar result of Geffert (*Proc. 7th DCFS*, Como, Italy, 23–37).

1 Introduction

Regular languages and finite automata are among the oldest and simplest topics in formal language theory. They have been intensively studied since the forties. Nevertheless, some important problems are still open. The most famous is the question of how many states are sufficient and necessary for two-way deterministic finite automata to simulate two-way nondeterministic finite automata [1,17].

Recently, there have been a new interest in automata theory; for a discussion, we refer to [10,20]. Many researchers have investigated various problems concerning descriptional complexity which studies the costs of description of languages by different formal systems. Here we focus on the deterministic and nondeterministic state complexity of complements, stars, and reversals of regular languages.

In 1997, at the 3rd Conference on Developments in Language Theory, Iwama at al. [11] 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 exactly α states for all integers n and α

* Research supported by the VEGA grant 2/6089/26.

M. Ito and M. Toyama (Eds.): DLT 2008, LNCS 5257, pp. 431–442, 2008.

satisfying that $n \leqslant \alpha \leqslant 2^n$. The question has also been considered in [12], where an integer Z with $n < Z < 2^n$ is called a "magic number" if no DFA of Z states can be simulated by any NFA of n states. In [13] it has been shown that there are no magic numbers, that is, appropriate automata have been described for all integers n and α. However, the constructions have used a growing alphabet of size $2^{n-1}+1$. Later, in [5], the size of the alphabet has been decreased to $n+2$, and finally, in [16], the result has been proved for a fixed four-letter alphabet. On the other hand, there are a lot of magic numbers in a unary case [6]. The problem remains open for binary and ternary alphabets.

A similar question for complements of regular languages has been examined in [15]. Using a growing alphabet of size 2^{n+1} it has been proved that all values in the range from $\log n$ to 2^n can be obtained as the nondeterministic state complexity of an n-state NFA language. Here we improve this result by showing that it still holds for a fixed five-letter alphabet. We also consider a binary case, and, as a bonus, we get an exponential number of so called "non-magic" values.

We next investigate the deterministic and nondeterministic state complexity of stars and reversals of regular languages. In all cases, we show that the whole range of complexities up to the known upper bounds can be obtain. To prove the results on state complexity we use growing alphabets. In the nondeterministic case, a binary alphabet is enough to describe appropriate automata.

To conclude this section let us mention some other related works. Magic numbers for symmetric difference NFAs have been studied by Zijl [22]. In [9], it has been shown that the deterministic and nondeterministic state complexity of union and intersection of regular languages may reach each value from 1 up to the upper bounds mn or $m+n+1$. Similar results for the nonterminal complexity of some operations on context-free languages have been recently obtained by Dassow and Stiebe [4].

2 Preliminaries

In this section, we give some basic definitions, notations, and preliminary results used throughout the paper. For further details, we refer to [18,19].

Let Σ be a finite alphabet and Σ^* the set of all strings over the alphabet Σ including the empty string ε. The length of a string w is denoted by $|w|$. A language is any subset of Σ^*. The complement of a language L is denoted by L^c, its star by L^*, and it reversal by L^R. We denote the cardinality of a finite set A by $|A|$ and its power-set by 2^A.

A *deterministic finite automaton* (DFA) is a 5-tuple $M = (Q, \Sigma, \delta, q_0, F)$, where Q is a finite set of states, Σ is a finite input alphabet, δ is the transition function that maps $Q \times \Sigma$ to Q, q_0 is the initial state, $q_0 \in Q$, and F is the set of accepting states, $F \subseteq Q$. In this paper, all DFAs are assumed to be complete, that is, the next state $\delta(q, a)$ is defined for each state q in Q and each symbol a in Σ. The transition function δ is extended to a function from $Q \times \Sigma^*$ to Q in a natural way. A string w in Σ^* is accepted by the DFA M if the state $\delta(q_0, w)$ is an accepting state of the DFA M. The *language accepted by* the DFA M, denoted $L(M)$, is the set of strings $\{w \in \Sigma^* \mid \delta(q_0, w) \in F\}$.

A *nondeterministic finite automaton* (NFA) is a 5-tuple $M = (Q, \Sigma, \delta, q_0, F)$, where Q, Σ, q_0 and F are defined in the same way as for a DFA, and δ is the nondeterministic transition function that maps $Q \times \Sigma$ to 2^Q. The transition function can be naturally extended to the domain $Q \times \Sigma^*$. A string w in Σ^* is accepted by the NFA M if the set $\delta(q_0, w)$ contains an accepting state of the NFA M. The *language accepted by* the NFA M is the set of strings $L(M) = \{w \in \Sigma^* \mid \delta(q_0, w) \cap F \neq \varnothing\}$.

Two automata are said to be *equivalent* if they accept the same language. A DFA (an NFA) M is called *minimal* if all DFAs (all NFAs, respectively) that are equivalent to M have at least as many states as M. It is well-known that a DFA $M = (Q, \Sigma, \delta, q_0, F)$ is minimal if *(i)* all its states are reachable from the initial state, and *(ii)* no two its different states are equivalent (states p and q are said to be equivalent if for all strings w in Σ^*, the state $\delta(p, w)$ is accepting iff the state $\delta(q, w)$ is accepting). Each regular language has a unique minimal DFA, up to isomorphism. However, the same result does not hold for NFAs.

The *(deterministic) state complexity* of a regular language is the number of states in its minimal DFA. The *nondeterministic state complexity* of a regular language is defined as the number of states in a minimal NFA accepting this language. A regular language with deterministic (nondeterministic) state complexity n is called an n-state DFA language (an n-state NFA language, respectively).

Every nondeterministic finite automaton $M = (Q, \Sigma, \delta, q_0, F)$ can be converted to an equivalent deterministic finite automaton $M' = (2^Q, \Sigma, \delta', q_0', F')$ using an algorithm known as the "subset construction" in the following way. Every state of the DFA M' is a subset of the state set Q. The initial state of the DFA M' is the set $\{q_0\}$. The transition function δ' is defined by $\delta'(R, a) = \bigcup_{r \in R} \delta(r, a)$ for each state R in 2^Q and each symbol a in Σ. A state R in 2^Q is an accepting state of the DFA M' if it contains at least one accepting state of the NFA M. The DFA M' need not be minimal since some states may be unreachable or equivalent. Sometimes, also NFAs with a set of initial states are considered. In such a case, the subset construction starts with this set being the initial state of an equivalent DFA.

To prove that an NFA is minimal we use a fooling-set lower-bound technique [2,3,7]. After defining a fooling set, we recall the lemma from [2] describing this lower-bound technique.

Definition 1. *A set of pairs of strings $\{(x_i, y_i) \mid i = 1, 2, \ldots, n\}$ is said to be a fooling set for a regular language L if for every i and j in $\{1, 2, \ldots, n\}$,*

(1) the string $x_i y_i$ is in the language L, and

(2) if $i \neq j$, then at least one of the strings $x_i y_j$ and $x_j y_i$ is not in L.

Lemma 1 (Birget [2]). *Let a set of pairs of strings $\{(x_i, y_i) \mid i = 1, 2, \ldots, n\}$ be a fooling set for a regular language L. Then every NFA for the language L needs at least n states.* $\square$

3 Complements

We start with the complements of regular languages. In the deterministic case, there is not much to say. The state complexity of a language and its complement is the same since to get a DFA for the complement we can simply exchange the accepting and the rejecting states in a DFA for the given language. The nondeterministic case is completely different. Given an n-state NFA we can apply the subset construction, and then exchange the accepting and the rejecting states, which gives an upper bound 2^n on the size of an NFA for the complement. This upper bound is known to be tight [17,3], and can be reached by the complement of a binary regular language [14].

Here we deal with the question of what values can be reached as the size of a minimal NFA accepting the complement of an n-state NFA language. In [15] it has been shown that all values from $\log n$ to 2^n can be reached, however, appropriate automata have been defined over a growing alphabet of size 2^{n+1}. In this section, we prove that this result still holds for a fixed five-letter alphabet. For each α with $\log n \leqslant \alpha \leqslant 2^n$, we describe a minimal n-state NFA M with a five-letter input alphabet such that every minimal NFA for the complement of the language $L(M)$ has exactly α states. In the second part of this section, we study a binary case, and show that here the whole range of complexities from $3 \log n$ to $n + 2^{n/3}$ can be obtained. As a bonus, we get an exponential number of so called non-magic values, which improves a similar result of Geffert [5].

The first two lemmata solve special cases of $\alpha = n$ and $\alpha = 2^n$. The next one has been recently proved in [16].

Lemma 2 ([15]). *For every $n \geqslant 1$, there exists a minimal binary NFA M of n states such that every minimal NFA for the complement of the language $L(M)$ has n states.* □

Lemma 3 ([14]). *For every $n \geqslant 1$, there exists a minimal binary NFA M of n states such that every minimal NFA for the complement of the language $L(M)$ has 2^n states.* □

Lemma 4 ([16], Theorem 1). *For all integers n and α with $n < \alpha < 2^n$, there exists a minimal NFA of n states with a four-letter input alphabet whose equivalent minimal DFA has exactly α states.* □

We use the automata from the lemma above to prove the next result which shows that the nondeterministic state complexity of the complement of an n-state NFA language over a five-letter alphabet may reach an arbitrary value from $n + 1$ to $2^n - 1$.

Lemma 5. *For all integers n and α with $n < \alpha < 2^n$, there exists a minimal NFA M of n states with a five-letter input alphabet such that every minimal NFA for the complement of the language $L(M)$ has α states.*

Proof. Let $n < \alpha < 2^n$. Then there is an integer k such that $1 \leqslant k \leqslant n - 1$ and $n - k + 2^k \leqslant \alpha < n - (k + 1) + 2^{k+1}$. It follows that $\alpha = n - (k + 1) + 2^k + m$, where m is an integer such that $1 \leqslant m < 2^k$.

Let $C = C_{n,k,m} = (Q, \{a,b,c,d\}, \delta_C, q_0, \{k\})$, where $Q = \{0, 1, \ldots, n-1\}$, be the n-state NFA from Lemma 4 whose minimal DFA has α states.

Now, let $M = M_{n,k,m} = (Q, \{a,b,c,d,f\}, \delta, q_0, \{k\})$ be an n-state NFA obtained from the NFA C by adding transitions on a new symbol f so that by f, state i with $0 \leqslant i \leqslant k-1$ goes to $\{i+1\}$, state k goes to $\{0, 1, \ldots, k\}$, and each other state goes to the empty set.

Let M' be the DFA obtained from the NFA M by the subset construction. It can be shown that the DFA M' has α reachable states. After exchanging the accepting and the rejecting states we get a DFA of the same number of states for the language $L(M)^c$. To prove the lemma it is sufficient to show that every NFA for the language $L(M)^c$ needs at least α states. This can be shown by describing a fooling set for the language $L(M)^c$ of size α. □

As a corollary of Lemmata 2, 3, and 5, and taking into account that $(L^c)^c = L$, we get the following result.

Theorem 1. *For all integers n and α with $\log n \leqslant \alpha \leqslant 2^n$, there exists a minimal nondeterministic finite automaton M of n states with a five-letter input alphabet such that every minimal nondeterministic finite automaton for the complement of the language $L(M)$ has exactly α states.* □

The second part of this section is devoted to the nondeterministic state complexity of the complements of binary regular languages. The first lemma deals with values from $n+4$ up to $2^{\lfloor n/3 \rfloor} - 1$, the second one covers the remaining cases.

Lemma 6. *For all integers n and α with $n + 4 \leqslant \alpha < n + 2^{\lfloor n/3 \rfloor}$, there exists a minimal binary NFA M of n states such that every minimal NFA for the complement of the language $L(M)$ has α states.*

Proof. Let $n+4 \leqslant \alpha < n+2^{\lfloor n/3 \rfloor}$ and let $k = \lfloor n/3 \rfloor$. Then α can be expressed as $\alpha = n + \sum_{i=0}^{k-1} c_i \cdot 2^i$, where $c_i \in \{0,1\}$ for $i = 0, 1, \ldots, k-1$. Denote by $m = \max\{i \mid c_i = 1\}$ and $\ell = |\{i > 0 \mid c_i = 1\}|$. Since $\alpha \geqslant n + 4$, we have $m \geqslant 2$.

Define an n-state NFA $M = (Q, \{a,b\}, \delta, p_1, \{1\})$, where $Q = \{p_1, p_2, \ldots, p_k\} \cup \{s_1, s_2, \ldots, s_k\} \cup \{1, 2, \ldots, n-2k\}$, and δ is defined as follows (see Fig. 1). If $1 \leqslant i < k$ and $c_i = 0$, then $\delta(p_i, a) = \{s_i\}$, $\delta(s_i, a) = \{p_{i+1}\}$, and $\delta(p_i, b) = \delta(s_i, b) = \varnothing$. If $1 \leqslant i < k$ and $c_i = 1$, then $\delta(p_i, a) = \{p_{i+1}\}$, $\delta(p_i, b) = \{s_i\}$, $\delta(s_i, a) = \{s_i, i\}$, and $\delta(s_i, b) = \{s_i\}$. Next, $\delta(p_k, a) = \{s_k\}$, $\delta(p_k, b) = \{\ell, \ell-1\}$, $\delta(s_k, a) = \{n-2k\}$, and $\delta(s_k, b) = \varnothing$ if $c_0 = 0$ and $\delta(s_k, b) = \{\ell+1, \ell\}$ if $c_0 = 1$. Finally, $\delta(q, a) = \delta(q, b) = \{q-1\}$ if $2 \leqslant q \leqslant n-2k$, $\delta(1, a) = \{s_m\} \cup \{1, 2, \ldots, m+1\}$, and $\delta(1, b) = \varnothing$.

Notice that there is a chain of a's going from state p_1 to state 1, which goes through all p_i's, those s_i's with $c_i = 0$, and states $n-2k, n-2k-1, \ldots, 2, 1$. The length of this chain is $n-1-\ell$, i.e., the string $a^{n-1-\ell}$ is in $L(M)$. Next, for all i with $c_i = 1$, all strings with an a in the i-th position from the end are accepted by M from state s_i, and no string in b^* is accepted from s_i.

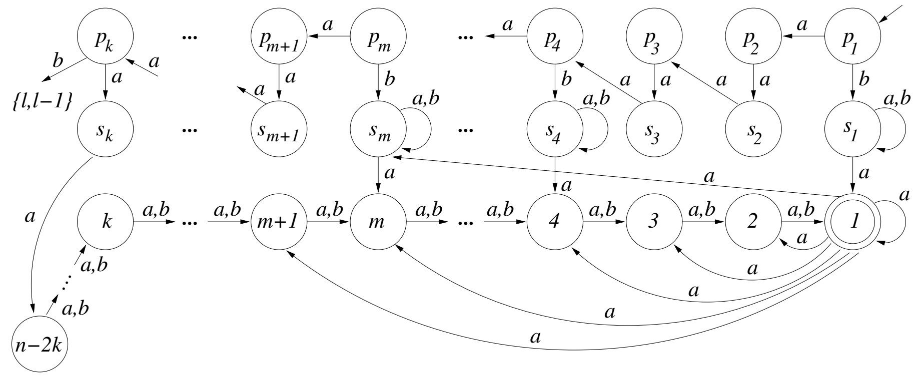

Fig. 1. The nondeterministic finite automaton M

It can be shown that the NFA M is minimal, the DFA M' obtained from the NFA M by the subset construction has α reachable states, and there is a fooling set for the language $L(M)^c$ of size α. $\square$

Lemma 7. *For all integers n and α with $n+1 \leqslant \alpha \leqslant 2n$, there exists a minimal binary NFA M of n states such that every minimal NFA for the complement of the language $L(M)$ has α states.* $\square$

By Lemmata 2, 6, 7, and the fact that $(L^c)^c = L$, we have the next result.

Theorem 2. *For all integers n and α with $3 \log n \leqslant \alpha < n + 2^{\lfloor n/3 \rfloor}$, there exists a minimal binary NFA M of n states such that every minimal NFA for the complement of the language $L(M)$ has exactly α states.* $\square$

As a corollary, we get an exponential number of non-magic values in a binary case, which improves the current number $2^{\Omega(n^{1/3} \ln^{2/3} n)}$ obtained by Geffert [5] using binary bounded languages.

Corollary 1. *For every $n \geqslant 1$, all values from n to $n + 2^{\lfloor n/3 \rfloor}$ are non-magic in a binary case, that is, for each integer α with $n \leqslant \alpha < n + 2^{\lfloor n/3 \rfloor}$, there exists a minimal binary NFA of n states whose equivalent minimal DFA has α states.*

Proof. Consider a binary NFA M described in Lemmata 2, 6, 7, for a given α. The DFA obtained from this NFA by the subset construction has α reachable sets. These sets must be inequivalent because otherwise we would have a smaller DFA for the language $L(M)$, and so, also a smaller DFA for the language $L(M)^c$. However, every NFA for the language $L(M)$ needs at least α states, a contradiction. Thus the minimal DFA for the language $L(M)$ has α states as desired. $\square$

4 Stars

This section deals with the deterministic and nondeterministic state complexity of stars of regular languages.

The upper bound on the state complexity of star operation is known to be $\frac{3}{4}2^n$ [21]. In the first part of this section, we show that each value from 1 to this upper bound can be reached as the state complexity of the star of an n-state DFA language. With an upper bound $n+1$, we prove a similar result for the nondeterministic state complexity of stars in the second part of this section. To get the result in the deterministic case we use a growing alphabet. In the nondeterministic case, a binary alphabet is enough to describe appropriate automata.

Let us start with recalling binary languages that reach the upper bound on the state complexity of star operation. Let $k \geqslant 2$ and let A_k be the binary k-state DFA depicted in Fig. 2. The following result has been shown by Yu, Zhuang and Salomaa [21].

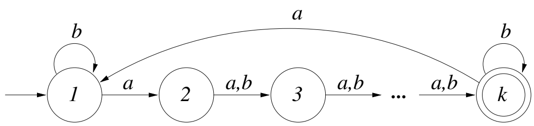

Fig. 2. The deterministic finite automaton A_k

Lemma 8 ([21]). *For every $k \geqslant 2$, the minimal DFA for the language $L(A_k)^*$ has $\frac{3}{4}2^k$ states.* □

Using automata A_k described above we prove the following lemma.

Lemma 9. *For all integers n and k with $2 \leqslant k \leqslant n$, there exists a minimal DFA $B_{n,k}$ of n states with a four-letter input alphabet such that the minimal DFA for the language $L(B_{n,k})^*$ has $n-k+\frac{3}{4}2^k$ states.*

Proof. If $k=n$, then take the DFA A_n from Lemma 8. Let $2 \leqslant k \leqslant n-1$ and let $\Sigma = \{a,b,c,d\}$.

Let us construct an n-state DFA $B_{n,k}$ with the input alphabet Σ from the k-state DFA A_k by adding new states $k+1, k+2, \ldots, n$, which go to itself by a,b,c except for state $k+1$ which goes to state 1 by a,b,c. Each of the states in $\{1,2,\ldots,k\}$ goes to state $k+1$ by c and to state $k+2$ by d. By d, state n goes to state 1, and state q with $k+1 \leqslant q \leqslant n-1$ to state $q+1$. The DFA $B_{n,k}$ is shown in Fig. 3 and is minimal since no two of its states are equivalent. If $k=n-1$, then the DFA $B_{n,k}$ is defined over the alphabet $\{a,b,c\}$.

Construct an NFA B' be for the language $L(B_{n,k})^*$ from the DFA $B_{n,k}$ by adding a new initial (and accepting) state q_0 which goes to state 2 by a, to state 1 by b, to state $k+1$ by c, and to state $k+2$ by d. Next, add transitions by a and by b from state $k-1$ to state 1.

Let B'' be the DFA obtained from the NFA B' by the subset construction. The DFA B'' has $n-k+\frac{3}{4}2^k$ reachable and pairwise inequivalent states, and the lemma follows. □

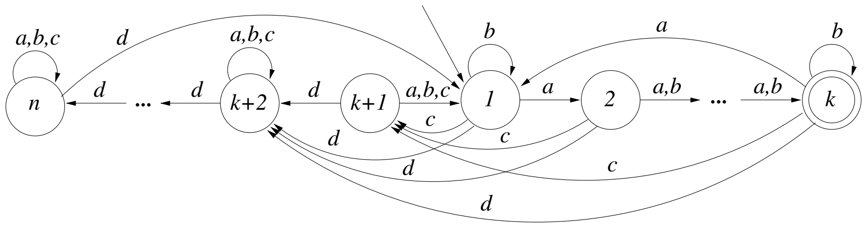

Fig. 3. The deterministic finite automaton $B_{n,k}$

Using automata $B_{n,k}$ we prove the following result showing that the state complexity of the star of an n-state DFA language may be arbitrary from $n+1$ to $\frac{3}{4}2^n$.

Lemma 10. *For all n and α with $n+1 \leqslant \alpha \leqslant \frac{3}{4}2^n$, there is a minimal DFA M of n states such that the minimal DFA for the language $L(M)^*$ has α states.*

Proof. If $\alpha = n - k + \frac{3}{4}2^k$, where $2 \leqslant k \leqslant n$, then take the n-state DFA $B_{n,k}$ from Lemma 9. Otherwise, let k be an integer such that $n - k + \frac{3}{4}2^k < \alpha < n - (k+1) + \frac{3}{4}2^{k+1}$. Then $\alpha = n - k + \frac{3}{4}2^k + m$ for some integer m with $1 \leqslant m \leqslant 2^{k-1} + 2^{k-2} - 2$.

Let $S_1, S_2, \ldots, S_\ell$, where $\ell = 2^{k-1} + 2^{k-2} - 2$, be all subsets of $\{1, 2, \ldots, k-1\}$ and all subsets $\{1, k\} \cup T$ with $T \subseteq \{2, 3, \ldots, k-1\}$, except for the emptyset and the set $\{1, 2, \ldots, k\}$, ordered in such a way that $S_1 = \{1\}$, and the sets of a smaller cardinality precede the sets with a larger cardinality. Now let $S_1, S_2, \ldots, S_m$ be the first m sets in the sequence.

Construct the DFA $M = M_{n,k,m}$ from the DFA $B_{n,k}$ by adding transitions on m new symbols $f_1, f_2, \ldots, f_m$ so that by symbol f_i $(1 \leqslant i \leqslant m)$, each state q in S_i goes to itself, and each state q in $\{1, 2, \ldots, n\} \setminus S_i$ goes to state $k+1$.

Let M' be an NFA for the language $L(M)^*$ obtained from the DFA M by adding a new initial (and accepting) state q_0 as in Lemma 9. By f_i $(1 \leqslant i \leqslant m)$, state q_0 goes to state 1 if $1 \in S_i$, and to state $k+1$ if $1 \notin S_i$. If the accepting state k is in S_i, then we add the transition by f_i from state k to state 1.

Let M'' be the DFA obtained from the NFA M' by the subset construction. The DFA M'' has $n - k + \frac{3}{4}2^k + m$ reachable and pairwise inequivalent states, which proves the lemma. □

The next lemma shows that sometimes even less than n states are sufficient to accept the star of an n-state DFA language. To describe appropriate automata it uses unary or binary alphabets.

Lemma 11. *For all integers n and k with and $1 \leqslant k \leqslant n$, there exists a minimal binary DFA M of n states such that the minimal DFA for the language $L(M)^*$ has k states.* □

Let us summarize the above results in the following theorem.

Theorem 3. *For all integers n and α with either $1 = n \leqslant \alpha \leqslant 2$, or $n \geqslant 2$ and $1 \leqslant \alpha \leqslant \frac{3}{4}2^n$, there exists a minimal DFA M of n states with a 2^n-letter input alphabet such that the minimal DFA for the star of the language $L(M)$ has exactly α states.* □

The upper bound on the nondeterministic state complexity of stars of n-state NFA languages is known to be $n+1$ [8]. The next theorem shows that each value from 1 to $n+1$ can be reached as the nondeterministic state complexity of the star of an n-state binary NFA language.

Theorem 4. *The nondeterministic state complexity of the star of each 1-state NFA language is 1. If $n \geqslant 2$, then for every k with $1 \leqslant k \leqslant n+1$, there exists a minimal NFA M of n states with a binary input alphabet such that every minimal NFA for the star of the language $L(M)$ has exactly k states.* □

5 Reversals

This section studies the deterministic and nondeterministic state complexity of reversals of regular languages.

If a regular language is accepted by an n-state DFA, then an n-state NFA for its reversal can be obtained from this DFA by interchanging the initial and the accepting states, and by reversing all transitions. By applying the subset construction to this NFA, we get a DFA for the reversal of at most 2^n states. Since the reversal of the reversal of a language is the same language, the lower bound on the size of the minimal DFA for the reversal of an n-state DFA language is $\log n$ (whenever $n \geqslant 3$; note that the reversal of an 1-state DFA language is the same language). In this section, we show that each value from $\log n$ to 2^n can be reached as the state complexity of the reversal of an n-state DFA language. In the second part of this section, we deal with the nondeterministic state complexity of reversals.

We start with the following lemma showing that all values from n to $2n$ can be reached as the state complexity of the reversal of an n-state DFA binary language.

Lemma 12. *For all integers n and α with $2 \leqslant n \leqslant \alpha \leqslant 2n$, there exists a minimal binary DFA A of n states such that the minimal DFA for the language $L(A)^R$ has α states.* □

The next lemma describes an n-state DFA language over an $(n-1)$-letter alphabet such that the state complexity of its reversal is $2n+1$. We use this automaton later in our constructions.

Lemma 13. *Let $n \geqslant 3$ and let $\Sigma = \{a_1, \ldots, a_{n-1}\}$ be an $(n-1)$-letter alphabet. There exists a minimal DFA B of n states with the input alphabet Σ such that the minimal DFA for the language $L(B)^R$ has $2n+1$ states.*

Proof. Define an n-state DFA $B = (Q, \Sigma, \delta, n, \{1\})$, where $Q = \{1, 2, \ldots, n\}$, and for all $q = 1, 2, \ldots, n$ and all $i = 1, 2, \ldots, n-1$, $\delta(i+1, a_i) = i$ and $\delta(q, a_i) = n$ if $q \neq i+1$, that is, by a_i, state $i+1$ goes to state i and each other state goes to state n. The DFA B is minimal since if $1 \leqslant i < j \leqslant n$, then the string $a_{i-1}a_{i-2}\cdots a_1$ is accepted by the DFA B from state i but not from state j.

Let B' be the NFA for the language $L(B)^R$ obtained from the DFA B by interchanging the accepting and the rejecting state and by reversing all transitions. Let B'' be the DFA obtained from the NFA B' by the subset construction. The DFA B'' has $2n+1$ reachable and pairwise inequivalent states, which proves the lemma. □

The next lemma deals with the case, when α is between $2n+2$ and 2^n, and uses a growing alphabet of size $n + \lfloor \alpha/2 \rfloor$ to describe appropriate automata.

Lemma 14. *For all integers n and α with $n \geqslant 3$ and $2n+2 \leqslant \alpha \leqslant 2^n$, there exists a minimal DFA C of n states such that the minimal DFA for the language $L(C)^R$ has α states.*

Proof. Let $\alpha = 2n + 1 + m$, where $1 \leqslant m \leqslant 2^n - 2n - 1$.

Let $k = \lfloor m/2 \rfloor$ and let $\Sigma_m = \{a_1, a_2, \ldots, a_{n-1}, b_1, b_2, \ldots, b_k\}$ if m is even, and $\Sigma_m = \{a_1, a_2, \ldots, a_{n-1}, b_1, b_2, \ldots, b_k, c\}$ if m is odd.

Let $Q = \{1, 2, \ldots, n\}$ and $T = \{2, 3, \ldots, n\}$. Now take all subsets of Q with cardinality more than 1, and order them in a sequence

$$S_1, Q \setminus S_1, S_2, Q \setminus S_2, \ldots, S_k, Q \setminus S_k, \ldots, S_{2^{n-1}-n-1}, Q \setminus S_{2^{n-1}-n-1},$$

(that is, each odd set of size at least two is followed by its complement in Q).

Define an n-state DFA $C = (Q, \Sigma_m, \delta, n, \{1\})$, in which for all $i = 1, \ldots, n-1$, the transitions by symbol a_i are the same as in the DFA B described in the proof of Lemma 13. Next, for all $j = 1, 2, \ldots, k$, by symbol b_j, each state in S_j goes to state 1, and each state in $Q \setminus S_j$ goes to state n. If m is odd, then, moreover, by symbol c, state 1 goes to state n, and each other state goes to state 1.

Let C' be the NFA for the language $L(C)^R$ obtained from the DFA C by interchanging the accepting and the rejecting state and by reversing all transitions. Let C'' be the DFA obtained from the NFA C' by the subset construction. The DFA C'' has $2n + 1 + m$ reachable and pairwise inequivalent states, which completes the proof of the lemma. □

As a corollary of the three lemmata above and using the fact that $(L^R)^R = L$ we get the following result.

Theorem 5. *For all integers n and α with $n \geqslant 3$ and $\log n \leqslant \alpha \leqslant 2^n$, there exists a minimal DFA M of n states with a 2^n-letter input alphabet such that the minimal DFA for the reversal of the language $L(M)$ has exactly α states. The minimal DFA for the reversal of a 2-state DFA language may have 2, 3, or 4 states, and the reversal of a 1-state DFA language is a 1-state DFA language.* □

We now turn our attention to the nondeterministic state complexity of reversals of regular languages represented by NFAs. The reversal of each 1-state NFA language is the same language. For $n \geqslant 2$, the upper bound on the nondeterministic state complexity of an n-state NFA language is known to be $n+1$ [8], and can be reached by the reversal of a binary language [14]. By the reversal of this binary language, in the case of $n \geqslant 3$, the lower bound $n-1$ is reached. The reversal of the n-state NFA language $\{w \in \{a,b\} \mid |w| \equiv 0 \mod n\}$ is the same language. Thus, the nondeterministic state complexity of a 2-state NFA language is 2 or 3, and for $n \geqslant 3$, we get the following result.

Theorem 6. *Let $n \geqslant 3$. Then the nondeterministic state complexity of the reversal of an n-state NFA binary language is either $n-1$, or n, or $n+1$.* □

6 Conclusions

We have investigated the deterministic and nondeterministic state complexity of complements, stars, and reversals of regular languages. In all cases, we have shown that the whole ranges of complexities up to the known upper bounds can be obtained. Our results are summarized in the following tables (where $[r \mathinner{..} s]$ denotes the set of all integers α with $r \leqslant \alpha \leqslant s$).

	State Complexity		Alphabet Size
L^c	$\{n\}$	trivial	arbitrary
L^*	$[1 \mathinner{..} \frac{3}{4}2^n]$	Theorem 3	2^n
L^R	$[\log n \mathinner{..} 2^n]$	Theorem 5	2^n

	Nondeterministic State Complexity		Alphabet Size
L^c	$[\log n \mathinner{..} 2^n]$	Theorem 1	5
L^*	$[1 \mathinner{..} n+1]$	Theorem 4	2
L^R	$\{n-1, n, n+1\}$	Theorem 6	2

To prove the results on nondeterministic state complexity we have used a fixed five-letter alphabet in the case of complements, and a binary alphabet in the case of stars and reversals. The results on the state complexity of stars and reversals have been shown for a growing alphabet. Whether or not they still hold for a fixed alphabet remains open. We also have proved some partial results on complements in a binary case, and, as a corollary, we have obtained exponentially many "non-magic" numbers, which improves a similar result of Geffert [5].

References

1. Berman, P., Lingas, A.: On the complexity of regular languages in terms of finite automata. Technical Report 304, Polish Academy of Sciences (1977)
2. Birget, J.C.: Intersection and union of regular languages and state complexity. Inform. Process. Lett. 43, 185–190 (1992)

3. Birget, J.C.: Partial orders on words, minimal elements of regular languages, and state complexity. Theoret. Comput. Sci. 119, 267–291 (1993)
4. Dassow, J., Stiebe, R.: Nonterminal complexity of some operations on context-free languages. In: Geffert, V., Pighizzini, G. (eds.) 9th International Workshop on Descriptional Complexity of Formal Systems, pp. 162–169. P. J. Šafárik University of Košice, Slovakia (2007)
5. Geffert, V. (Non)determinism and the size of one-way finite automata. In: Mereghetti, C., Palano, B., Pighizzini, G., Wotschke, D. (eds.) 7th International Workshop on Descriptional Complexity of Formal Systems, pp. 23–37. University of Milano, Italy (2005)
6. Geffert, V.: Magic numbers in the state hierarchy of finite automata. In: Královič, R., Urzyczyn, P. (eds.) MFCS 2006. LNCS, vol. 4162, pp. 412–423. Springer, Heidelberg (2006)
7. Glaister, I., Shallit, J.: A lower bound technique for the size of nondeterministic finite automata. Inform. Process. Lett. 59, 75–77 (1996)
8. Holzer, M., Kutrib, M.: Nondeterministic descriptional complexity of regular languages. Internat. J. Found. Comput. Sci. 14, 1087–1102 (2003)
9. Hricko, M., Jirásková, G., Szabari, A.: Union and intersection of regular languages and descriptional complexity. In: Mereghetti, C., Palano, B., Pighizzini, G., Wotschke, D. (eds.) 7th International Workshop on Descriptional Complexity of Formal Systems, pp. 170–181. University of Milano, Italy (2005)
10. Hromkovič, J.: Descriptional complexity of finite automata: Concepts and open problems. J. Autom. Lang. Comb. 7, 519–531 (2002)
11. 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); Preliminary version In:Bozapalidis, S. (ed.) 3rd International Conference on Developments in Language Theory. Aristotle University of Thessaloniki (1997)
12. Iwama, K., Matsuura, A., Paterson, M.: A family of NFAs which need $2^n - \alpha$ deterministic states. Theoret. Comput. Sci. 301, 451–462 (2003)
13. Jirásková, G.: Note on minimal finite automata. In: Sgall, J., Pultr, A., Kolman, P. (eds.) MFCS 2001. LNCS, vol. 2136, pp. 421–431. Springer, Heidelberg (2001)
14. Jirásková, G.: State complexity of some operations on binary regular languages. Theoret. Comput. Sci. 330, 287–298 (2005)
15. Jirásek, J., Jirásková, G., Szabari, A.: State complexity of concatenation and complementation. Internat. J. Found. Comput. Sci. 16, 511–529 (2005)
16. Jirásek, J., Jirásková, G., Szabari, A.: Deterministic blow-ups of minimal nondeterministic finite automata over a fixed alphabet. In: Harju, T., Karhumäki, J., Lepistö, A. (eds.) DLT 2007. LNCS, vol. 4588, pp. 254–265. Springer, Heidelberg (2007)
17. Sakoda, W.J., Sipser, M.: Nondeterminism and the size of two-way finite automata. In: 10th Annual ACM Symposium on Theory of Computing, San Diego, California, USA, pp. 275–286 (1978)
18. Sipser, M.: Introduction to the theory of computation. PWS Publishing Company, Boston (1997)
19. Yu, S.: Regular languages. In: Rozenberg, G., Salomaa, A. (eds.) Handbook of Formal Languages, ch. 2, vol. I, pp. 41–110. Springer, Heidelberg (1997)
20. Yu, S.: A renaissance of automata theory? Bull. Eur. Assoc. Theor. Comput. Sci. 72, 270–272 (2000)
21. Yu, S., Zhuang, Q., Salomaa, K.: The state complexity of some basic operations on regular languages. Theoret. Comput. Sci. 125, 315–328 (1994)
22. Zijl, L.: Magic numbers for symmetric difference NFAs. Internat. J. Found. Comput. Sci. 16, 1027–1038 (2005)