Operational Complexity in Subregular Classes

Michal Hospodár[ID] and Galina Jirásková[(✉)][ID]

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

Abstract. The state complexity of a regular operation is a function that assigns the maximal state complexity of the language resulting from the operation to the sizes of deterministic finite automata recognizing the operands of the operation. We study the state complexity of intersection, union, concatenation, star, and reversal on the classes of combinational, singleton, finitely generated left ideal, symmetric definite, star, comet, two-sided comet, ordered, star-free, and power-separating languages. We get the exact state complexities in all cases. The complexity of all operations on combinational languages is given by a constant function. The state complexity of the considered operations on singleton languages is $\min\{m,n\}$, $m+n-3$, $m+n-2$, $n-1$, and n, respectively, and on finitely generated left ideals, it is $mn-2$, $mn-2$, $m+n-1$, $n+1$, and $2^{n-2}+2$. The state complexity of concatenation, star, and reversal is $m2^n - 2^{n-1} - m + 1$, n, 2^n and $m2^{n-1} - 2^{n-2} + 1$, $n+1$, $2^{n-1}+1$ for star and symmetric definite languages, respectively. We also show that the complexity of reversal on ordered and power-separating languages is $2^n - 1$, which proves that the lower bound for star-free languages given by [Brzozowski, Liu, Int. J. Found. Comput. Sci. 23, 1261–1276, 2012] is tight. In all the remaining cases, the complexity is the same as for regular languages. Except for reversal on finitely generated left ideals and ordered languages, all our witnesses are described over a fixed alphabet.

1 Introduction

The state complexity of a regular operation is the number of states that is sufficient and necessary in the worst case for a deterministic finite automaton (DFA) to accept the language resulting from the operation, considered as a function of the sizes of DFAs for operands. The tight upper bounds on the complexity of union, intersection, concatenation, star, and reversal were presented by Maslov [12], Yu, Zhuang and Salomaa [17], and Leiss [11], and to describe witness languages, a binary alphabet was used.

If operands have some special properties, then the complexity of an operation may be significantly smaller. For example, the state complexity of star on prefix-free languages is n, while it is $\frac{3}{4}2^n$ in the regular case. On the other hand, the complexity of intersection, union, concatenation, and star on star-free languages is the same as in the regular case [5]. This led to the investigation of operational complexity in several subclasses of regular languages. Operations on unary languages were studied by Yu et al. [17] and Pighizzini and Shallit [15], and on finite languages by Câmpeanu et al. [6]. The classes of prefix- and suffix-free languages were examined by Han et al. [7,8], and the classes of bifix-, factor-, and subword-free, star-free, ideal, and closed languages by Brzozowski et al. [2–5], Except for reversal on star-free languages, the state complexity of basic regular operations in each of the above mentioned classes was determined. Bordihn, Holzer, and Kutrib [1] considered also some other subclasses, including combinational, finitely generated left ideal, symmetric definite, star, comet, two-sided comet, ordered, and power-separating languages, and investigated the complexity of the conversion of nondeterministic finite automata (NFAs) to deterministic ones. The closure properties of these classes, as well as nondeterministic operational complexity in them, were studied by Olejár et al. [9,13].

Here we continue this research and study the state complexity of basic regular operations in several classes from [1]. For each considered operation and each considered class, we get a tight upper bound on its complexity. To get the upper bounds, we examine minimal DFAs for languages in considered classes. We show that every minimal DFA for a finitely generated left ideal must have two states that can be distinguished only by the empty string. This gives upper bounds for union, intersection, and reversal on finitely generated left ideals. We also show that the set of all non-final states in a minimal DFA for a power-separating language cannot be reachable in the reversed automaton. This gives an upper bound $2^n - 1$ for reversal on power-separating languages, and shows that the same lower bound for star-free languages [5, Theorem 7] is tight since every star-free language is power-separating. To get upper bounds for concatenation, we carefully inspect the reachable and distinguishable states in the subset automaton of an NFA for the concatenation of given languages. Finally, the star of a star language is the same language, and if a given language is symmetric definite, then its star differs from it only in the empty string. The corresponding upper bounds n and $n + 1$ for the star operation follow. To get lower bounds, we sometimes use witnesses known from the literature, and just prove that they belong to a considered class. However, most often, we describe appropriate witnesses so that the corresponding product automata for union and intersection, or subset automata of NFAs for concatenation, star, or reversal have a desired number of reachable and pairwise distinguishable states.

2   Preliminaries

We assume that the reader is familiar with the basic notions in formal languages and automata theory, and for details, we refer to [16].

Let Σ be a finite non-empty alphabet of symbols. Then Σ^* denotes the set of all strings over Σ including the empty string ε. For a finite set S, the notation $|S|$ denotes the size of S and 2^S denotes the power set of S. For two non-negative integers i, j, the set $\{i, i+1, \ldots, j\}$ is denoted by $[i, j]$.

A *nondeterministic finite automaton with multiple initial states* (MNFA) is a quintuple $A = (Q, \Sigma, \cdot, I, F)$ where Q is a finite non-empty set of states, Σ is a finite non-empty input alphabet, $\cdot: Q \times \Sigma \to 2^Q$ is the transition function, $I \subseteq Q$ is the set of initial states, and $F \subseteq Q$ is the set of final states. The transition function can be extended to the domain $2^Q \times \Sigma^*$ in the natural way. For states p, q and a symbol a, we sometimes write (p, a, q) whenever $q \in p \cdot a$. The language accepted by A is the set of strings $L(A) = \{w \in \Sigma^* \mid I \cdot w \cap F \neq \emptyset\}$.

The *reverse* of an MNFA $A = (Q, \Sigma, \cdot, I, F)$ is the MNFA $A^R = (Q, \Sigma, \cdot^R, F, I)$ where $q \cdot^R a = \{p \mid q \in p \cdot a\}$. We say that a subset S of Q is *reachable* in A if there exists a string w such that $S = I \cdot w$. A subset S is *co-reachable* in A if it is reachable in A^R. We usually omit $\cdot$ and write just juxtaposition qw instead of $q \cdot w$. If $S \subseteq Q$, then $Sw = \{qw \mid q \in S\}$ and $wS = \{q \mid qw \in S\}$.

An MNFA is a *nondeterministic finite automaton* (NFA) if $|I| = 1$. An NFA is a *deterministic finite automaton* (DFA) if $|qa| = 1$ for each state q and each symbol a. We usually write $pa = q$ instead of $pa = \{q\}$. A non-final state q of a DFA is called *dead* if $qa = q$ for each symbol a. A DFA is *minimal* if all its states are reachable and pairwise distinguishable. The *state complexity of a regular language* L, $\mathrm{sc}(L)$, is the number of states in a minimal DFA recognizing L. The *state complexity of a k-ary regular operation* f is the function from $\mathbb{N}^k$ to $\mathbb{N}$ defined as $(n_1, n_2, \ldots, n_k) \mapsto \max\{\mathrm{sc}(f(L_1, L_2, \ldots, L_k)) \mid \mathrm{sc}(L_i) \leq n_i \text{ for each } i.\}$

Every MNFA $A = (Q, \Sigma, \cdot, I, F)$ can be converted to an equivalent deterministic automaton $\mathcal{D}(A) = (2^Q, \Sigma, \cdot, I, \{S \in 2^Q \mid S \cap F \neq \emptyset\})$. The DFA $\mathcal{D}(A)$ is called the *subset automaton* of A.

Finally, we recall the definitions of language classes considered in this paper. A language $L \subseteq \Sigma$ is:

- *combinational* (class abbreviation CB): if $L = \Sigma^* H$ for $H \subseteq \Sigma$;
- *singleton* (SGL): if it consists of one string;
- *finitely generated left ideal* (FGLID): if $L = \Sigma^* H$ for some finite language H;
- *left ideal* (LID): if $L = \Sigma^* L$;
- *symmetric definite* (SYDEF): if $L = G\Sigma^* H$ for some regular languages G, H;
- *star* (STAR): if $L = L^*$;
- *comet* (COM): if $L = G^* H$ for some regular languages G, H with $G \notin \{\emptyset, \{\varepsilon\}\}$;
- *two-sided comet* (2COM): if $L = EG^* H$ for some regular languages E, G, H with $G \notin \{\emptyset, \{\varepsilon\}\}$;
- *ordered* (ORD): if it is accepted by a (possibly non-minimal) DFA with ordered states such that $p \preceq q$ implies $p \cdot \sigma \preceq q \cdot \sigma$ for each symbol σ;
- *star-free* (STFR): if L is constructible from finite languages, concatenation, union, and complementation (equivalently, if L has an aperiodic DFA);
- *power-separating* (PSEP): if for every w in Σ^* there exists an integer k such that $\bigcup_{i \geq k} \{w^i\} \subseteq L$ or $\bigcup_{i \geq k} \{w^i\} \subseteq L^c$.

We have CB $\subsetneq$ FGLID $\subsetneq$ LID $\subsetneq$ SYDEF, STAR $\setminus \{\{\varepsilon\}\} \subsetneq$ COM $\subsetneq$ 2COM, and SGL $\subsetneq$ ORD $\subsetneq$ STFR $\subsetneq$ PSEP [1,13].

3   Combinational and Singleton Languages

We start with two simple subregular classes: the class of combinational languages and the class of singleton languages. The following two theorems present our results on the state complexity of basic operations in these two classes.

Theorem 1. *Let K and L be combinational languages over Σ. Then we have $\mathrm{sc}(K), \mathrm{sc}(L) \leq 2$ and $\mathrm{sc}(K \cap L), \mathrm{sc}(K \cup L) \leq 2$, $\mathrm{sc}(KL) \leq 3$, $\mathrm{sc}(L^*) \leq 2$, and $\mathrm{sc}(L^R) \leq 3$. All these upper bounds are tight if $|\Sigma| \geq 2$ (star, reversal) or if $|\Sigma| \geq 1$ (intersection, union, concatenation).*

Proof. Let $K = \Sigma^*G$ and $L = \Sigma^*H$ with $G, H \subseteq \Sigma$. If H is empty, then L is empty. Otherwise, L is accepted by a two-state DFA $(\{1,2\}, \Sigma, \cdot, 1, \{2\})$ with $1a = 2a = 1$ if $a \notin H$ and $1a = 2a = 2$ if $a \in H$. We only prove the result for star. Each string in Σ^* either ends with a symbol in H, or ends with a symbol in $\Sigma \backslash H$, or is empty. Since $L = L^+$ we have $L^* = \{\varepsilon\} \cup L = (\Sigma^*(\Sigma \backslash H))^c$, so L^* is the complement of a combinational language. The upper bound 2 is met by $(a + b)^*a$.   □

Theorem 2. *Let K and L be singleton languages accepted by DFAs with m and n states, respectively. Then $\mathrm{sc}(K \cap L) \leq \min\{m, n\}$, $\mathrm{sc}(K \cup L) \leq m + n - 3$, $\mathrm{sc}(KL) \leq m + n - 2$, $\mathrm{sc}(L^*) \leq n - 1$, and $\mathrm{sc}(L^R) \leq n$. All these bounds are tight, with witnesses described over an alphabet of size at least 1 (intersection, concatenation, reversal) or 2 (union, star).*

Proof. We only prove the result for star. The minimal DFA for a singleton language $\{v\}$ has $|v| + 2$ states, including the dead state. Next, the language $\{v\}^*$ is accepted by a DFA with $|v| + 1$ states, possibly including the dead state. This gives the upper bound $n - 1$, which is met by the language $\{a^{n-2}\}$ over the binary alphabet $\{a, b\}$.   □

4   Intersection and Union

In this section, we examine the intersection and union operations in subregular classes. Recall that the state complexity of these two operations is mn with binary witnesses [12,17]. Our aim is to show that the complexity of these operations is $mn - 2$ in the class of finitely generated left ideals and mn in the remaining classes. Recall that if a DFA for a language L has a unique final state which is the initial one, then $L = L^*$, so L is a star language. To get the upper bound in class FGLID, we use the following property.

Lemma 3. *In the minimal DFA for a finitely generated left ideal different from $\emptyset$ and Σ^*, there exist two states that are distinguishable only by the empty string.*

Proof. Let A be a minimal DFA for a finitely generated left ideal L with the initial state 1. Assume, to get a contradiction, that any two distinct states of A can be distinguished by a non-empty string. Since $L \notin \{\emptyset, \Sigma^*\}$, there is a symbol a
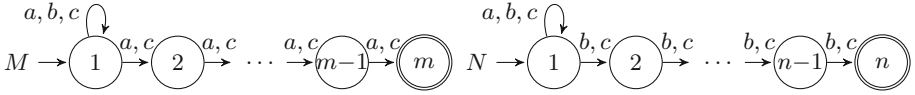
Fig. 1. Finitely generated left ideal witnesses for union $(mn - 2)$.

such that $1a \neq 1$. Now, the states 1 and $1a$ are distinguishable by a non-empty string w_1. We must have $1w_1 \notin F$ and $1aw_1 \in F$, because $1w_1 \in F$ and $1aw_1 \notin F$ would mean that $w_1 \in L$ and $aw_1 \notin L$, so L would not be a left ideal. The resulting states are again distinguishable by a non-empty string w_2, and again, we must have $1w_1w_2 \notin F$ and $1aw_1w_2 \in F$. Inductively, the states $1w_1w_2\cdots w_{i-1}$ and $1aw_1w_2\cdots w_{i-1}$ are distinguishable by a non-empty string w_i, and we must have $1w_1w_2\cdots w_{i-1}w_i \notin F$ and $1aw_1w_2\cdots w_{i-1}w_i \in F$. Since the number of states is finite, there exist j, k with $j < k$ such that $1aw_1w_2\cdots w_k = 1aw_1w_2\cdots w_j$. This means that $aw_1w_2\cdots w_j(w_{j+1}w_{j+2}\cdots w_k)^* \subseteq L$, and on the other hand we have $w_1w_2\cdots w_j(w_{j+1}w_{j+2}\cdots w_k)^* \subseteq L^c$. It follows that the minimal generator $L \setminus \Sigma L$ of the left ideal L is infinite, a contradiction. $\square$

Theorem 4. *Let K and L be finitely generated left ideals over Σ accepted by an m-state and n-state DFA, respectively. Then $\mathrm{sc}(K \cap L), \mathrm{sc}(K \cup L) \leq mn - 2$. This upper bound is tight for union if $|\Sigma| \geq 3$, and for intersection if $|\Sigma| \geq 10$.*

Proof. Let A and B be minimal DFAs for K and L, respectively. By Lemma 3, there exist states p, p' of A and q, q' of B that are distinguished only by the empty string. It follows that $pa = p'a$ and $qa = q'a$ for each input symbol a. Then, in the product automaton $A \times B$, the pairs $(p, q), (p, q'), (p', q), (p', q')$ are sent to the same state by each input symbol, so, they may be distinguished only by ε. This gives the upper bound $mn - 2$ for union and intersection.

We can prove that this upper bound for union and intersection is met by languages accepted by NFAs from Figs. 1 and 3, respectively; notice that equivalent DFAs shown in Figs. 2 and 4 have m and n states as well. $\square$

Theorem 5. *The state complexity of intersection and union on the classes of symmetric definite, ordered, power-separating, and star languages is mn.*

Proof. The upper bound mn is met by the intersection of binary left ideal languages $K = \{w \in \{a, b\}^* \mid |w|_a \geq m - 1\}$ and $L = \{w \in \{a, b\}^* \mid |w|_b \geq n - 1\}$
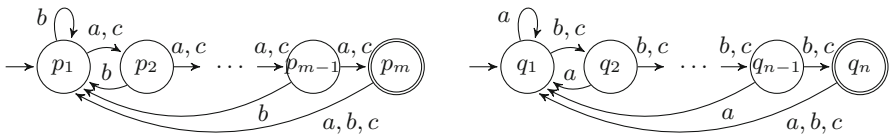

Fig. 2. The reachable parts of the subset automata $\mathcal{D}(M)$ and $\mathcal{D}(N)$ of NFAs M and N from Fig. 1; we have $p_i = [1, i]$ and $q_j = [1, j]$.
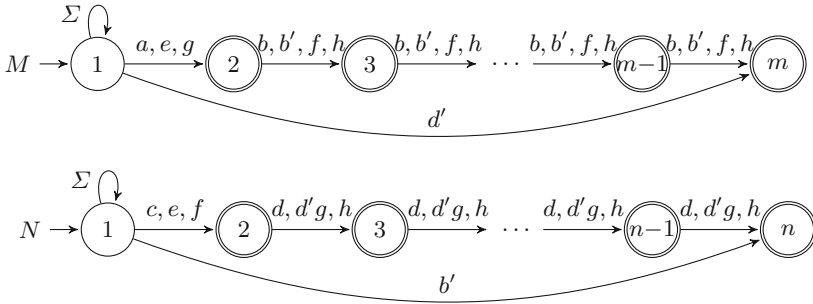
Fig. 3. Finitely generated left ideal witnesses for intersection $(mn-2)$.

[3, Theorem 7], and by the union of quaternary left ideals from [3, Theorem 8]. Every left ideal is symmetric definite. Next, the languages K and L are ordered (so also power-separating), and these classes are closed under complementation. This gives the complexity mn with binary witnesses on ORD and PSEP. Finally, to get star witnesses, we consider the minimal DFAs, the first of which counts the number of a's modulo m, and the second the number of b's modulo n. □

Remark 6. Since STAR $\subseteq$ COM $\subseteq$ 2COM, the previous theorem gives complexity mn for intersection and union on comet and two-sided comet languages. □

5   Concatenation

Now we examine the concatenation operation, the state complexity of which is $m2^n - 2^{n-1}$ with binary witnesses [12]. We first consider the concatenation operation on finitely generated left ideal, symmetric definite, and star languages, and show that the resulting complexities are always smaller than the regular
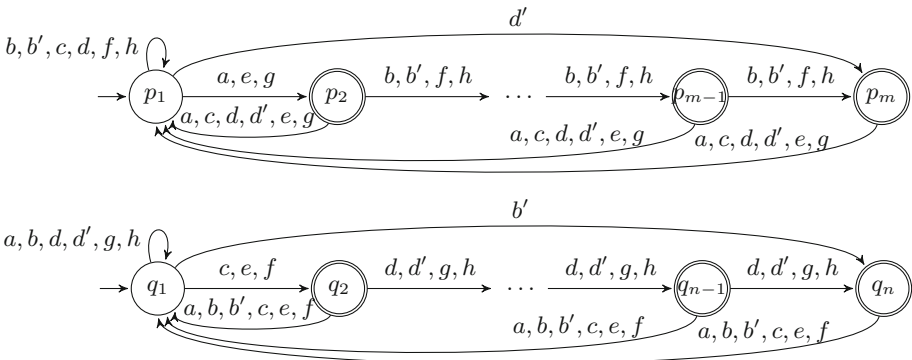

Fig. 4. The reachable parts of the subset automata $\mathcal{D}(M)$ and $\mathcal{D}(N)$ of NFAs M and N from Fig. 3; we have $p_i = \{1, i\}$ and $q_j = \{1, j\}$.

one. On the other hand, in the remaining classes, the complexity is the same as in the regular case. To get an upper bound for SYDEF, we use the following observation, cf. [14, Theorem 9.2].

Lemma 7. *Every minimal DFA for a symmetric definite language contains a state $\bar{q}$ such that each accepting computation goes through $\bar{q}$ and a loop on every input symbol can be added in $\bar{q}$ without changing the language.*   □

Remark 8. As shown in [3, Theorem 9], the upper bound on the complexity of concatenation on left ideals is $m + n - 1$. This upper bound is shown to be met by unary languages a^*a^{m-1} and a^*a^{n-1} in the proof of [3, Theorem 9]. These unary languages are finitely generated left ideals. This gives the same complexity of concatenation on FGLID.   □

Theorem 9. *Let $K, L \subseteq \Sigma^*$ be accepted by an m-state and n-state DFA, respectively. If K and L are symmetric definite, then $\mathrm{sc}(KL) \leq m2^{n-1} - 2^{n-2} + 1$. If K and L are star languages, then $\mathrm{sc}(KL) \leq m2^n - 2^{n-1} - m + 1$. Both upper bounds are tight if $|\Sigma| \geq 3$.*

Proof. Let $A = (Q_A, \Sigma, \cdot_A, s_A, F_A)$ and $B = (Q_B, \Sigma, \cdot_B, s_B, F_B)$ be the minimal DFAs for K and L, respectively.

(a) By Lemma 7, there is a state $\bar{q}$ of B such that every accepting computation of B goes through the state $\bar{q}$ and, moreover, we can add a loop on every input symbol in $\bar{q}$ without changing the language; denote the resulting NFA by B'. Construct the MNFA N for KL from automata A and B' by adding the transition (q, a, s_B) whenever $qa \in F_A$, the initial set is $\{s_A\}$ if $s_A \notin F_A$ and $\{s_A, s_B\}$ otherwise, and the set of final states is F_B. Each reachable subset of the subset automaton of N is of the form $\{p\} \cup S$ with $p \in Q_A$ and $S \subseteq Q_B$; let us represent it by the pair (p, S). Recall that $p \in F_A$ implies $s_B \in S$ for each reachable state (p, S).

If $\bar{q}$ is the initial state of B, then L is a left ideal, and the complexity of KL is at most $m + n - 1$; cf. [3, Theorem 9]. If $\bar{q}$ is final, then in the corresponding subset automaton, each state $(p, S \cup \{\bar{q}\})$ is equivalent to the state $(p, \{\bar{q}\})$. This gives the upper bound $m2^{n-1} - 2^{n-2} + 1$. Finally, let $\bar{q}$ be neither initial nor final. Let Q_1 be the set of states in Q_B reachable without going through $\bar{q}$ and let $Q_2 = Q_B \backslash (Q_1 \cup \{\bar{q}\})$. Let $S_1 \subseteq Q_1$ and $S_2 \subseteq Q_2$. Every accepting computation of B goes through $\bar{q}$, and $\bar{q}$ has a loop on every input symbol in B'. Let us show that each state $(p, S_1 \cup \{\bar{q}\} \cup S_2)$ is equivalent to $(p, \{\bar{q}\} \cup S_2)$. It is enough to shown that each string accepted in B' from a state in S_1 is also accepted from a state in $\{\bar{q}\} \cup S_2$. Let a string w be accepted from S_1. Then the computation on w must go through $\bar{q}$, so $w = uv$ where u leads a state in S_1 to the state $\bar{q}$ and v is accepted from $\bar{q}$. Since $\bar{q}$ has a loop on each symbol in B', the string uv is also accepted from $\bar{q}$. Hence w is accepted from $(p, \{\bar{q}\} \cup S_2)$. This gives an upper bound $(m - 1)(2^{|Q_1|} + 2^{|Q_2|}) + 2^{|Q_1|-1} + 2^{|Q_2|} + 1 \leq (m - 1)(2^{n-1}) + 2^{n-2} + 1$.

For tightness, consider the DFAs A and B from Fig. 5. The automaton A recognizes a left ideal since, after adding a loop on each input symbol in its initial state and performing determinization and minimization, we get the isomorphic
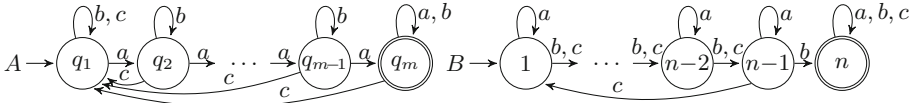
Fig. 5. Symmetric definite witnesses for concatenation ($m2^{n-1} - 2^{n-2} + 1$).

automaton. The automaton B recognizes a right ideal since it has a loop on each input symbol in its unique final state. Thus both automata recognize symmetric definite languages, and we can show that they meet the desired upper bound.

(b) If K and L are star languages, then $\varepsilon \in K$, so the initial state of the DFA for K is final. It follows that the MNFA for KL has two initial states, and so the initial state of the corresponding subset automaton is $(s_A, \{s_B\})$. This means that the states $(p, \emptyset)$ are unreachable. This gives the desired upper bound.

For tightness, consider the DFAs A and B from Fig. 6. Since the initial state is the unique final state, both automata recognize star languages.

Construct the MNFA M for $L(A)L(B)$ from A and B by adding the transitions $(q_1, b, 1)$ and $(q_m, a, 1)$ and making the state q_1 non-final; the set of initial states is $\{q_1, 1\}$. Notice that in the reversed automaton M^R, each singleton $\{j\}$ is reached from $\{1\}$ by a string in cb^*, and each singleton $\{q_i\}$ is reached by a string in ba^*. It follows that for each state p of M, there is a string w_p such that $\{p\}$ is reached from $\{1\}$ by w_p in M^R. It follows that the string w_p^R is accepted by M from and only from the state p. This means that all states of the subset automaton $\mathcal{D}(M)$ are pairwise distinguishable. Let us denote the (reachable) state $\{q_i\} \cup S$ with $S \subseteq [1, n]$ by (i, S), and let us show that the set of reachable states is $\{(1, S) \mid S \subseteq [1, n]$ with $1 \in S\} \cup \{(i, S) \mid i = 2, 3, \ldots, m$ and $S \subseteq [1, n]$ with $|S| \geq 1\}$. The initial state is $(1, \{1\})$ and it is sent to $(i, \{j\})$ by $a^{i-1}b^{j-1}$. Next, each (i, S) with $1 \in S$ is reached from $(m, S \setminus \{1\})$ by a^i and each (i, S) with $i \neq 1$, $S \neq \emptyset$, and $\min S = j \geq 2$ is reached from a state $(i, b^{j-1}S)$ with $1 \in b^{j-1}S$ by b^{j-1}.   □

Remark 10. If A and B are the binary DFAs shown in Fig. 7, cf. [12], then we have $\mathrm{sc}(L(A)L(B)) = m2^n - 2^{n-1}$ since in the NFA for concatenation, each singleton set is co-reachable, and the reachability of the desired number of states is shown similarly as in Theorem 9(b). Since $L(A) = b^*L(A)$ and $L(B) = a^*L(B)$, automata A and B recognize comet languages. Moreover we have COM $\subseteq$ 2COM, so the state complexity of concatenation on comets and two-sided comets is the same as in the class of regular languages.   □
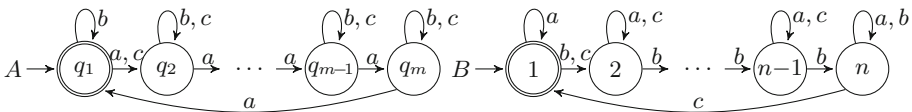

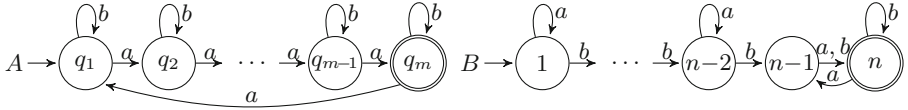**Fig. 6.** Star witnesses for concatenation ($m2^n - 2^{n-1} - m + 1$).

Fig. 7. Comet witnesses for concatenation $(m2^n - 2^{n-1})$.

Remark 11. The star-free witness languages that meet the regular upper bound for concatenation are described in [5, Theorem 2], and they are recognized by the DFAs shown in Fig. 8. Notice that both these DFAs are ordered. Since we have ORD $\subseteq$ STFR $\subseteq$ PSEP, the state complexity of concatenation on ordered and power-separating languages is $m2^n - 2^{n-1}$. □

6   Star and Reversal

In this section, we consider the star and reversal operations. The state complexity of star is $\frac{3}{4}2^n$ with binary witnesses [17]. Our aim is to show that the complexity of star in our classes is always the same as in the regular case, except for classes of finitely generated left ideal, symmetric definite, and star languages, where it is $n + 1$, $n + 1$, and n, respectively. Recall that FGLID $\subseteq$ SYDEF.

Theorem 12. *Let $n \geq 3$. Let L be a symmetric definite language accepted by an n-state DFA. Then L^* is accepted by a DFA with $n + 1$ states. This upper bound is met by the finitely generated left ideal $L = (a + b)^* a^{n-1}$.*

Proof. If $L \in$ SYDEF then $L = G\Sigma^* H$. Hence $L^+ = LL^* = G\Sigma^* H(G\Sigma^* H)^* = G\Sigma^* (HG\Sigma^*)^* H = G\Sigma^* H = L$. It follows that $L^* = \{\varepsilon\} \cup L$. This gives the upper bound $n + 1$. The reader may verify that $L = (a + b)^* a^{n-1}$ is accepted by an n-state DFA and $\mathrm{sc}(\{\varepsilon\} \cup L) = n + 1$. □

Remark 13. If L is a star language, then $L = L^*$, so $\mathrm{sc}(L^*) = \mathrm{sc}(L)$. □

Remark 14. The regular witness for star [17, Theorem 3.3] is recognized by the DFA A from Fig. 9. Since $L(A) = b^* L(A)$, the language $L(A)$ is a comet, so also a two-sided comet. Thus the complexity of star on COM and 2COM is $\frac{3}{4}2^n$. □
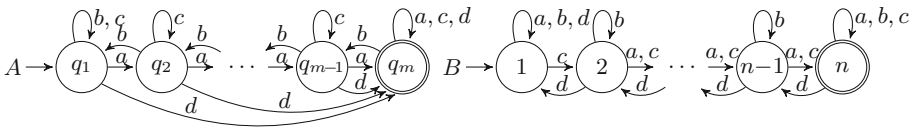

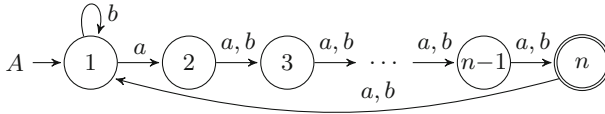**Fig. 8.** Ordered witnesses for concatenation $(m2^n - 2^{n-1})$.

Fig. 9. A comet witness for star $(\frac{3}{4}2^n)$.

Remark 15. The regular upper bound $\frac{3}{4}2^n$ is met by the quaternary star-free language [5, Theorem 6] which is accepted by the DFA shown in Fig. 10. Since this DFA is ordered, and every ordered language is power-separating, the state complexity of star on ORD and PSEP is the same as in the regular case.   □

We conclude our investigation with the reversal operation, the state complexity of which is 2^n with binary witnesses [11]. First, using Lemmas 3 and 7, we get tight upper bounds for finitely generated left ideals and symmetric definite languages. Then, by proving that the set of all non-final states of a DFA accepting a power-separating language cannot be co-reachable, we get the tight upper bound $2^n - 1$ for reversal on ordered, star-free, and power-separating languages, which shows that a lower bound for star-free languages [5, Theorem 7] is tight. Finally, we show that the complexity of reversal on star, comet, and two-sided comet languages is the same as in the regular case.

Theorem 16. *Let L be a finitely generated left ideal over Σ accepted by an n-state DFA. Then $\mathrm{sc}(L^R) \leq 2^{n-2} + 2$, and this bound is tight if $|\Sigma| \geq 2^{n-2} + 2$.*

Proof. Let A be the minimal DFA for L. By Lemma 3, A has two states q, q' with $qa = q'a$ for each input a, and such that exactly one of them, say q, is final. Since L is a left ideal, adding a loop on each symbol in the initial state of A does not change the language; denote the resulting NFA by N. Thus all final sets of the subset automaton of N^R are equivalent. Moreover, except for possible initial set $\{q\}$, each reachable set contains either both q, q', or none of them.

For tightness, let $\Sigma = \{a, b\} \cup \{c_S \mid S \subseteq [2, n-1]\}$. Consider an n-state NFA $N = ([1, n], \Sigma, \cdot, 1, \{n\})$ where $1 \cdot a = \{1, 2\}$, $1 \cdot b = \{1\}$ and $i \cdot b = \{i + 1\}$ if $2 \leq i \leq n - 1$, $1 \cdot c_S = \{1\}$ and $i \cdot c_S = \{n\}$ if $i \in S$, and all the remaining transitions go to the empty set. The NFA N recognizes a finitely generated left ideal since state 1 has a loop on each symbol and for any other transition (i, σ, j) we have $i < j$. Since the subset automaton of N has n reachable states, the language $L(N)$ is accepted by an n-state DFA. In the reverse N^R, the initial set
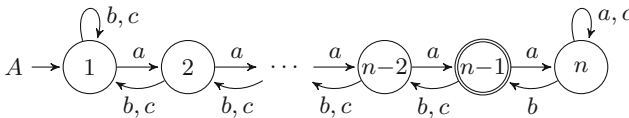

Fig. 10. An ordered witness for star; transitions (i, d, n) for each i are not shown $(\frac{3}{4}2^n)$.

is $\{n\}$, and it is sent to a subset S of $[2, n-1]$ by c_S. The set $\{1\}$ is reached from $\{2\}$ by a. For distinguishability, notice that $\{1\}$ is the unique final state. If $S, T \subseteq [2, n]$ and $s \in S \setminus T$, then $b^{s-2}a$ sends S to the final set $\{1\}$ and it sends T to the empty set. □

Theorem 17. *Let $L \subseteq \Sigma^*$ be a symmetric definite language accepted by an n-state DFA. Then $\mathrm{sc}(L^R) \leq 2^{n-1} + 1$, and this upper bound is tight if $|\Sigma| \geq 3$.*

Proof. We get an upper bound in a similar way as in the proof of Theorem 9(a). The ternary left ideal from [3, Theorem 11(2), Fig. 17] meets this bound. □

Theorem 18. *Let L be a power-separating language accepted by an n-state DFA. Then $\mathrm{sc}(L^R) \leq 2^n - 1$, and this bound is met by an ordered language over an alphabet of size $n - 1$.*

Proof. Let $A = (Q, \Sigma, \cdot, s, F)$ be a minimal DFA for a power-separating language L. Our aim is to show that the set $Q \setminus F$ of all non-final states of A cannot be co-reachable in A. Assume that $Q \setminus F$ is co-reachable by a string w. Since w sends F to $Q \setminus F$ in A^R, the string w must be rejected by A from each final state, and accepted from each non-final state, so there is no $k \geq 0$ such that $\bigcup_{i \geq k}\{w^i\} \subseteq L$ or $\bigcup_{i \geq k}\{w^i\} \subseteq L^c$. This means that L is not power-separating, a contradiction.

To get tightness, let $\Sigma = \{a, b\} \cup \{c_i \mid i = 3, 4, \ldots, n-1\}$. Consider the n-state DFA $A = ([1, n], \Sigma, \cdot, 1, F)$ where $F = \{i \in [1, n] \mid i \bmod 2 = 0\}$, and for each i in $[1, n]$ and each $j = 3, 4, \ldots, n-1$, let $i \cdot a = i+1$ if $i \leq n-1$ and $n \cdot a = n$, $i \cdot b = i-1$ if $i \geq 2$ and $1 \cdot b = 1$, $i \cdot c_j = i$ if $i \neq j$ and $j \cdot c_j = j-1$. Then A is ordered. It is shown in [5, Theorem 7] that every DFA for the reverse of $L(A)$ has at least $2^n - 1$ states. □

Remark 19. The DFA A from Fig. 11 differs from Šebej's witness for reversal meeting the upper bound 2^n [10, Theorem 5] only in the set of final states, and Šebej's proof works for this DFA as well. Since the initial state of A is a unique final state, A recognizes a star, so also comet and two-sided comet, language. □
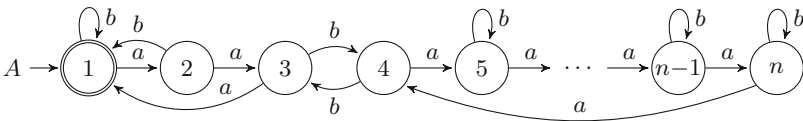

Fig. 11. A star witness for reversal (2^n).

7   Conclusion

We investigated the state complexity of basic regular operations in the classes of combinational, singleton, finitely generated left ideal, symmetric definite, star, comet, two-sided comet, ordered, and power-separating languages. Our results are summarized in Table 1 where the sizes of alphabets used to describe witnesses are given by numbers in parentheses. Except for the ordered and finitely generated left ideal witnesses for reversal, all the remaining witnesses are described over a fixed alphabet, and whenever a binary alphabet is used, it is always optimal. By showing that the upper bound for reversal on power-separating languages is $2^n - 1$, we proved that the same lower bound for reversal on star-free languages from [5, Theorem 7] is tight.

In [1], the classes of definite, generalized definite, locally testable, and strictly locally testable languages were also considered. We leave the operational complexity in these classes for future research. The optimality of alphabet sizes greater than two remains open as well.

Table 1. Operational complexity in subregular classes; we have $\bullet = m2^{n-1} - 2^{n-2} + 1$ and $\circ = m2^n - 2^{n-1} - m + 1$.

class	$K \cap L$	$K \cup L$	KL	L^*	L^R
CB Th. 1	2 (1)	2 (1)	3 (1)	2 (2)	3 (2)
SGL Th. 2	$\min\{m,n\}$ (1)	$m + n - 3$ (2)	$m + n - 2$ (1)	$n - 1$ (2)	n (1)
FGLID	$mn - 2$ (10)	$mn - 2$ (3)	$m + n - 1$ (1)	$n + 1$ (2)	$2^{n-2} + 2$
source		Th. 4	Re. 8	Th. 12	Th. 16
LID	mn (2)	mn (4)	$m + n - 1$ (1)	$n + 1$ (2)	$2^{n-1} + 1$ (2)
source		[3, Th. 8]	[3, Th. 9]	[3, Th. 10]	[3, Th. 11]
SYDEF	mn (2)	mn (4)	$\bullet$ (3)	$n + 1$ (2)	$2^{n-1} + 1$ (3)
source		Th. 5	Th. 9	Th. 12	Th. 17
STAR	mn (2)	mn (2)	$\circ$ (3)	n (1)	2^n (2)
source		Th. 5	Th. 9	Re. 13	Re. 19
COM, 2COM	mn (2)	mn (2)	$m2^n - 2^{n-1}$ (2)	$\frac{3}{4}2^n$ (2)	2^n (2)
source		Re. 6	Re. 10	Re. 14	Re. 19
ORD, PSEP	mn (2)	mn (2)	$m2^n - 2^{n-1}$ (4)	$\frac{3}{4}2^n$ (4)	$2^n - 1$ $(n-1)$
source		Th. 5	Re. 11	Re. 15	Th. 18
STFR	mn (2)	mn (2)	$m2^n - 2^{n-1}$ (4)	$\frac{3}{4}2^n$ (4)	$2^n - 1$ $(n-1)$
source		[5, Th. 1]	[5, Th. 2]	[5, Th. 6]	Th. 18
regular	mn (2)	mn (2)	$m2^n - 2^{n-1}$ (2)	$\frac{3}{4}2^n$ (2)	2^n (2)
source		[17, Th. 4.3]	[12]	[17, Th. 3.3]	[11, Prop. 2]

References

1. Bordihn, H., Holzer, M., Kutrib, M.: Determination of finite automata accepting subregular languages. Theor. Comput. Sci. **410**(35), 3209–3222 (2009). https://doi.org/10.1016/j.tcs.2009.05.019

2. Brzozowski, J., Jirásková, G., Li, B., Smith, J.: Quotient complexity of bifix-, factor-, and subword-free regular languages. Acta Cybern. **21**(4), 507–527 (2014). https://doi.org/10.14232/actacyb.21.4.2014.1

3. Brzozowski, J.A., Jirásková, G., Li, B.: Quotient complexity of ideal languages. Theor. Comput. Sci. **470**, 36–52 (2013). https://doi.org/10.1016/j.tcs.2012.10.055

4. Brzozowski, J., Jirásková, G., Zou, C.: Quotient complexity of closed languages. Theor. Comput. Sci. **54**(2), 277–292 (2013). https://doi.org/10.1007/s00224-013-9515-7

5. Brzozowski, J.A., Liu, B.: Quotient complexity of star-free languages. Int. J. Found. Comput. Sci. **23**(6), 1261–1276 (2012). https://doi.org/10.1142/S0129054112400515

6. Câmpeanu, C., Culik, K., Salomaa, K., Yu, S.: State complexity of basic operations on finite languages. In: Boldt, O., Jürgensen, H. (eds.) WIA 1999. LNCS, vol. 2214, pp. 60–70. Springer, Heidelberg (2001). https://doi.org/10.1007/3-540-45526-4_6

7. Han, Y.S., Salomaa, K.: State complexity of basic operations on suffix-free regular languages. Theor. Comput. Sci. **410**(27–29), 2537–2548 (2009). https://doi.org/10.1016/j.tcs.2008.12.054

8. Han, Y.S., Salomaa, K., Wood, D.: Operational state complexity of prefix-free regular languages. In: Ésik, Z., Fülöp, Z. (eds.) Automata, Formal Languages, and Related Topics, pp. 99–115. University of Szeged, Hungary (2009)

9. Hospodár, M., Mlynárčik, P., Olejár, V.: Operations on subregular languages and nondeterministic state complexity. In: Han, Y., Vaszil, G. (eds.) DCFS 2022. LNCS, vol. 13439, pp. 112–126. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-13257-5_9

10. Jirásková, G., Šebej, J.: Reversal of binary regular languages. Theor. Comput. Sci. **449**, 85–92 (2012). https://doi.org/10.1016/j.tcs.2012.05.008

11. Leiss, E.L.: Succinct representation of regular languages by Boolean automata. Theor. Comput. Sci. **13**, 323–330 (1981). https://doi.org/10.1016/S0304-3975(81)80005-9

12. Maslov, A.N.: Estimates of the number of states of finite automata. Soviet Math. Doklady **11**, 1373–1375 (1970)

13. Olejár, V., Szabari, A.: Closure properties of subregular languages under operations. Int. J. Found. Comput. Sci., online ready. https://doi.org/10.1142/S0129054123450016. Extended abstract in MCU 2022. LNCS, vol. 13419, pp. 126–142. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-13502-6_9

14. Paz, A., Peleg, B.: Ultimate-definite and symmetric-definite events and automata. J. ACM **12**(3), 399–410 (1965). https://doi.org/10.1145/321281.321292

15. Pighizzini, G., Shallit, J.: Unary language operations, state complexity and Jacobsthal's function. Int. J. Found. Comput. Sci. **13**(1), 145–159 (2002). https://doi.org/10.1142/S012905410200100X

16. Sipser, M.: Introduction to the theory of computation. Cengage Learning (2012)

17. Yu, S., Zhuang, Q., Salomaa, K.: The state complexities of some basic operations on regular languages. Theor. Comput. Sci. **125**(2), 315–328 (1994). https://doi.org/10.1016/0304-3975(92)00011-F