Syntactic Complexities of Some Classes of Star-Free Languages*

Janusz Brzozowski and Baiyu Li

David R. Cheriton School of Computer Science,
University of Waterloo, Waterloo, ON, Canada N2L 3G1
{brzozo,b5li}@uwaterloo.ca

Abstract. The syntactic complexity of a regular language is the cardinality of its syntactic semigroup. The syntactic complexity of a subclass of regular languages is the maximal syntactic complexity of languages in that subclass, taken as a function of the state complexity n of these languages. We study the syntactic complexity of three subclasses of star-free languages. We find tight upper bounds for languages accepted by monotonic, partially monotonic and "nearly monotonic" automata; all three of these classes are star-free. We conjecture that the bound for nearly monotonic languages is also a tight upper bound for star-free languages.

Keywords: finite automaton, monotonic, nearly monotonic, partially monotonic, star-free language, syntactic complexity, syntactic semigroup.

1 Introduction

Star-free languages are the smallest class containing the finite languages and closed under boolean operations and concatenation. In 1965, Schützenberger proved [19] that a language is star-free if and only if its syntactic monoid is *group-free*, that is, has only trivial subgroups. An equivalent condition is that the minimal deterministic automaton of a star-free language is *permutation-free*, that is, has only trivial permutations (cycles of length 1). Such automata are called *aperiodic,* and this is the term we use. Star-free languages were studied in detail in 1971 by McNaughton and Papert [15].

The *state complexity of a regular language* is the number of states in the minimal deterministic finite automaton (DFA) recognizing that language. State complexity of operations on languages has been studied quite extensively; for a survey of this topic and a list of references see [21]. An equivalent notion is that of *quotient complexity* [2], which is the number of left quotients of the language.

Quotient complexity is closely related to the Nerode equivalence [17]. Another well-known equivalence relation, the Myhill equivalence [16], defines the syntactic semigroup of a language and its *syntactic complexity*, which is the cardinality of the syntactic semigroup. It was pointed out in [5] that syntactic complexity can be very different for languages with the same quotient complexity.

* This work was supported by the Natural Sciences and Engineering Research Council of Canada under grant No. OGP0000871.

M. Kutrib, N. Moreira, and R. Reis (Eds.): DCFS 2012, LNCS 7386, pp. 117–129, 2012.

In contrast to state complexity, syntactic complexity has not received much attention. Suppose L is a regular language and has quotient complexity n. In 1970 Maslov [14] noted that n^n is a tight upper bound on the syntactic complexity of L. In 2003–2004 Holzer and König [9], and Krawetz, Lawrence and Shallit [12] studied the syntactic complexity of unary and binary languages. In 2010 Brzozowski and Ye [5] showed that, if L is any right ideal, then n^{n-1} is a tight upper bound on its syntactic complexity. They also proved that $n^{n-1} + (n-1)$ (respectively, $n^{n-2} + (n-2)2^{n-2} + 1$) is a lower bound if L a left (respectively, two-sided) ideal. In 2012 Brzozowski, Li and Ye [3] showed that n^{n-2} is a tight upper bound for prefix-free languages and that $(n-1)^{n-2} + (n-2)$ (respectively, $(n-1)^{n-3} + (n-2)^{n-3} + (n-3)2^{n-3}$ or $(n-1)^{n-3} + (n-3)2^{n-3} + 1$) is a lower bound for suffix-free (respectively, bifix-free or factor-free) languages.

Here we deal with star-free languages. It has been shown in 2011 by Brzozowski and Liu [4] that boolean operations, concatenation, star, and reversal in the class of star-free languages meet all the quotient complexity bounds of regular languages, with very few exceptions. Also, Kutrib, Holzer, and Meckel [10] proved in 2012 that in most cases exactly the same tight state complexity bounds are reached by operations on aperiodic nondeterministic finite automata (NFA's) as on general NFA's. In sharp contrast to this, the syntactic complexity of star-free languages appears to be much smaller than the n^n bound for regular languages. We derive tight upper bounds for three subclasses of star-free languages, the monotonic, partially monotonic, and nearly monotonic languages. We conjecture that the bound for star-free languages is the same as that for nearly monotonic languages.

The remainder of the paper is structured as follows. Our terminology and some basic facts are stated in Section 2. Aperiodic transformations are examined in Section 3. In Section 4, we study monotonic, partially monotonic, and nearly monotonic automata and languages. Section 5 concludes the paper.

2 Preliminaries

We assume the reader is familiar with basic theory of formal languages as in [18], for example. Let Σ be a non-empty finite alphabet and Σ^*, the free monoid generated by Σ. A *word* is any element of Σ^*, and the empty word is ε. The length of a word $w \in \Sigma^*$ is $|w|$. A *language* over Σ is any subset of Σ^*. For any languages K and L over Σ, we use the *boolean operations*: complement ($\overline{L}$) and union ($K \cup L$). The *product*, or *(con)catenation*, of K and L is $KL = \{w \in \Sigma^* \mid w = uv, u \in K, v \in L\}$; the *star* of L is $L^* = \bigcup_{i \geqslant 0} L^i$, and the *positive closure* of L is $L^+ = \bigcup_{i \geqslant 1} L^i$.

We call languages $\emptyset$, $\{\varepsilon\}$, and $\{a\}$ for any $a \in \Sigma$ the *basic* languages. *Regular* languages are the smallest class of languages constructed from the basic languages using boolean operations, product, and star. *Star-free* languages are the smallest class of languages constructed from the basic languages using only boolean operations and product.

A *deterministic finite automaton* (DFA) is a quintuple $\mathcal{A} = (Q, \Sigma, \delta, q_1, F)$, where Q is a finite, non-empty set of *states*, Σ is a finite non-empty *alphabet*,

$\delta : Q \times \Sigma \to Q$ is the *transition function*, $q_1 \in Q$ is the *initial state*, and $F \subseteq Q$ is the set of *final states*. We extend δ to $Q \times \Sigma^*$ in the usual way. The DFA $\mathcal{A}$ accepts a word $w \in \Sigma^*$ if $\delta(q_1, w) \in F$. The set of all words *accepted* by $\mathcal{A}$ is $L(\mathcal{A})$. Regular languages are exactly the languages accepted by DFA's. By the *language of a state* q of $\mathcal{A}$ we mean the language L_q accepted by the DFA $(Q, \Sigma, \delta, q, F)$. A state is *empty* if its language is empty.

An *incomplete deterministic finite automaton (IDFA)* is a quintuple $\mathcal{I} = (Q, \Sigma, \delta, q_1, F)$, where Q, Σ, q_1 and F are as in a DFA, and δ is a partial function. Every DFA is also an IDFA.

The *left quotient*, or simply *quotient,* of a language L by a word w is the language $w^{-1}L = \{x \in \Sigma^* \mid wx \in L\}$. The *Nerode equivalence* $\sim_L$ of any language L over Σ is defined as follows [17]: For all $x, y \in \Sigma^*$,

$$x \sim_L y \text{ if and only if } xv \in L \Leftrightarrow yv \in L, \text{ for all } v \in \Sigma^*.$$

Clearly, $x^{-1}L = y^{-1}L$ if and only if $x \sim_L y$. Thus each equivalence class of the Nerode equivalence corresponds to a distinct quotient of L.

Let L be a regular language. The *quotient DFA* of L is $\mathcal{A} = (Q, \Sigma, \delta, q_1, F)$, where $Q = \{w^{-1}L \mid w \in \Sigma^*\}$, $\delta(w^{-1}L, a) = (wa)^{-1}L$, $q_1 = \varepsilon^{-1}L = L$, and $F = \{w^{-1}L \mid \varepsilon \in w^{-1}L\}$. State $w^{-1}L$ of a quotient DFA is reachable from the initial state L by the word w. Also, the language of every state is distinct, since only distinct quotients are used as states. Thus every quotient DFA is minimal. The *quotient IDFA* of L is the quotient DFA of L after the empty state, if present, and all transitions incident to it are removed. The quotient IDFA is also minimal. If a regular language L has quotient IDFA $\mathcal{I}$, then the DFA $\mathcal{A}$ obtained by adding the empty state to $\mathcal{I}$, if necessary, is the quotient DFA of L. The two automata $\mathcal{A}$ and $\mathcal{I}$ accept the same language.

The number $\kappa(L)$ of distinct quotients of L is the *quotient complexity* of L. Since the quotient DFA of L is minimal, quotient complexity is the same as state complexity. The quotient viewpoint is often useful for deriving upper bounds, while the state approach may be more convenient for proving lower bounds.

The *Myhill equivalence* $\approx_L$ of L is defined as follows [16]: For all $x, y \in \Sigma^*$,

$$x \approx_L y \text{ if and only if } uxv \in L \Leftrightarrow uyv \in L \text{ for all } u, v \in \Sigma^*.$$

This equivalence is also known as the *syntactic congruence* of L. The quotient set $\Sigma^+/\approx_L$ of equivalence classes of the relation $\approx_L$ is a semigroup called the *syntactic semigroup* of L (which we denote by S_L), and $\Sigma^*/\approx_L$ is the *syntactic monoid* of L. The *syntactic complexity* $\sigma(L)$ of L is the cardinality of its syntactic semigroup. The *monoid complexity* $\mu(L)$ of L is the cardinality of its syntactic monoid. If the equivalence class containing ε is a singleton in the syntactic monoid, then $\sigma(L) = \mu(L) - 1$; otherwise, $\sigma(L) = \mu(L)$.

A *partial transformation* of a set Q is a partial mapping of Q into itself; we consider partial transformations of finite sets only, and we assume without loss of generality that $Q = \{1, 2, \ldots, n\}$. Let t be a partial transformation of Q. If t is defined for $i \in Q$, then it is the image of i under t; otherwise it is undefined and we write $it = \Box$. For convenience, we let $\Box t = \Box$. If X is a subset of Q, then

$Xt = \{it \mid i \in X\}$. The *composition* of two partial transformations t_1 and t_2 of Q is a partial transformation $t_1 \circ t_2$ such that $i(t_1 \circ t_2) = (it_1)t_2$ for all $i \in Q$. We usually drop the composition operator "$\circ$" and write t_1t_2 for short.

An arbitrary partial transformation can be written in the form

$$t = \begin{pmatrix} 1 & 2 & \cdots & n-1 & n \\ i_1 & i_2 & \cdots & i_{n-1} & i_n \end{pmatrix},$$

where $i_k = kt$ and $i_k \in Q \cup \{\Box\}$, for $k \in Q$. The *domain* of t is the set $\mathrm{dom}(t) = \{k \in Q \mid kt \neq \Box\}$. The *range* of t is the set $\mathrm{rng}(t) = \mathrm{dom}(t)t = \{kt \mid k \in Q \text{ and } kt \neq \Box\}$. When the domain is clear, we also write $t = [i_1, \dots, i_n]$.

A *(full) transformation* t of a set Q is a partial transformation such that $\mathrm{dom}(t) = Q$. The *identity* transformation maps each element to itself, that is, $it = i$ for $i = 1, \dots, n$. A transformation t is a *cycle* of length $k \geqslant 2$ if there exist pairwise distinct elements $i_1, \dots, i_k$ such that $i_1t = i_2, i_2t = i_3, \dots, i_{k-1}t = i_k$, $i_kt = i_1$, and $jt = j$ for all $j \not\in \{i_1, \dots, i_k\}$. Such a cycle is denoted by $(i_1, i_2, \dots, i_k)$. For $i < j$, a *transposition* is the cycle (i, j). A *singular* transformation, denoted by $\binom{i}{j}$, has $it = j$ and $ht = h$ for all $h \neq i$. A *constant* transformation, denoted by $\binom{Q}{j}$, has $it = j$ for all i. Let $\mathcal{T}_Q$ be the set of all transformations of Q, which is a semigroup under composition.

Let $\mathcal{A} = (Q, \Sigma, \delta, q_1, F)$ be a DFA. For each word $w \in \Sigma^+$, the transition function defines a transformation t_w of Q: for all $i \in Q$, $it_w \stackrel{\text{def}}{=} \delta(i, w)$. The set $T_{\mathcal{A}}$ of all such transformations by non-empty words forms a subsemigroup of $\mathcal{T}_Q$, called the *transition semigroup* of $\mathcal{A}$ [18]. Conversely, we can use a set $\{t_a \mid a \in \Sigma\}$ of transformations to define δ, and so the DFA $\mathcal{A}$. When the context is clear we simply write $a = t$, where t is a transformation of Q, to mean that the transformation performed by $a \in \Sigma$ is t. If $\mathcal{A}$ is the quotient DFA of L, then $T_{\mathcal{A}}$ is isomorphic to the syntactic semigroup S_L of L [15], and we represent elements of S_L by transformations in $T_{\mathcal{A}}$.

For any IDFA $\mathcal{I}$, each word $w \in \Sigma^*$ performs a partial transformation of Q. The set of all such partial transformations is the *transition semigroup* of $\mathcal{I}$. If $\mathcal{I}$ is the quotient IDFA of a language L, this semigroup is isomorphic to the transition semigroup of the quotient DFA of L, and hence also to the syntactic semigroup of L.

3 Aperiodic Transformations

A transformation is *aperiodic* if it contains no cycles of length greater than 1. A semigroup T of transformations is *aperiodic* if and only if it contains only aperiodic transformations. Thus a language L with quotient DFA $\mathcal{A}$ is star-free if and only if every transformation in $T_{\mathcal{A}}$ is aperiodic.

Let A_n be the set of all aperiodic transformations of Q. Each aperiodic transformation can be characterized by a forest of labeled rooted trees as follows. Consider, for example, the forest of Fig. 1 (a), where the roots are at the bottom. Convert this forest into a directed graph by adding a direction from each child

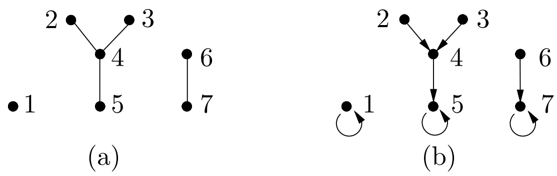

Fig. 1. Forests and transformations

to its parent and a self-loop to each root, as shown in Fig. 1 (b). This directed graph defines the transformation $[1,4,4,5,5,7,7]$ and such a transformation is aperiodic since the directed graph has no cycles of length greater than one. Thus there is a one-to-one correspondence between aperiodic transformations of a set of n elements and forests with n nodes.

Proposition 1. *There are* $(n+1)^{n-1}$ *aperiodic transformations of a set of* $n \geqslant 1$ *elements.*

Proof. By Cayley's theorem [6,20], there are $(n+1)^{n-1}$ labeled unrooted trees with $n+1$ nodes. If we fix one node, say node $n+1$, in each of these trees to be the root, then we have $(n+1)^{n-1}$ labeled trees rooted at $n+1$. Let T be any one of these trees, and let $v_1, \dots, v_m$ be the parents of $n+1$ in T. By removing the root $n+1$ from each such rooted tree, we get a labeled forest F with n nodes formed by m rooted trees, where $v_1, \dots, v_m$ are the roots. The forest F is unique since T is a unique tree rooted at $n+1$. Then we get a unique aperiodic transformation of $\{1, \dots, n\}$ by adding self-loops on $v_1, \dots, v_m$.

All labeled directed forests with n nodes can be obtained uniquely from some rooted tree with $n+1$ nodes by deleting the root. Hence there are $(n+1)^{n-1}$ labeled forests with n nodes, and that many aperiodic transformations of Q. □

Since the quotient DFA of a star-free language can perform only aperiodic transformations, we have

Corollary 2. *For* $n \geqslant 1$, *the syntactic complexity* $\sigma(L)$ *of a star-free language* L *with* n *quotients satisfies* $\sigma(L) \leqslant (n+1)^{n-1}$.

The bound of Corollary 2 is our first upper bound on the syntactic complexity of a star-free language with n quotients, but this bound is not tight in general because the set A_n *is not a semigroup* for $n \geqslant 3$. For example, if $a = [1,3,1]$ and $b = [2,2,1]$, then $ab = [2,1,2]$, which contains the cycle $(1,2)$. Hence our task is to find the size of the largest semigroup contained in A_n.

First, let us consider small values of n:

1. If $n = 1$, the only two languages, $\emptyset$ and Σ^*, are both star-free, since $\Sigma^* = \overline{\emptyset}$. Here $\sigma(L) = 1$, for both languages, the bound $2^0 = 1$ of Corollary 2 holds and it is tight.
2. If $n = 2$, $|A_2| = 3$. The only unary languages are ε and $\overline{\varepsilon} = aa^*$, and $\sigma(L) = 1$ for both. For $\Sigma = \{a,b\}$, one verifies that $\sigma(L) \leqslant 2$, and $\Sigma^* a \Sigma^*$ meets this bound. If $\Sigma = \{a,b,c\}$, then $L = \Sigma^* a \overline{\Sigma^* b \Sigma^*}$ has $\sigma(L) = 3$.

In summary, for $n = 1$ and 2, the bound of Corollary 2 is tight for $|\Sigma| = 1$ and $|\Sigma| = 3$, respectively.

We say that two aperiodic transformations a and b *conflict* if ab or ba contains a cycle; then (a, b) is called a *conflicting pair*. When $n = 3$, $|A_3| = 4^2 = 16$. The transformations $a_0 = [1, 2, 3]$, $a_1 = [1, 1, 1]$, $a_2 = [2, 2, 2]$, $a_3 = [3, 3, 3]$ cannot create any conflict. Hence we consider only the remaining 12 transformations.

Let $b_1 = [1, 1, 3]$, $b_2 = [1, 2, 1]$, $b_3 = [1, 2, 2]$, $b_4 = [1, 3, 3]$, $b_5 = [2, 2, 3]$, and $b_6 = [3, 2, 3]$. Each of them has only one conflict. There are also two *conflicting triples* (b_1, b_3, b_6) and (b_2, b_4, b_5), since $b_1b_3b_6$ and $b_2b_4b_5$ both contains a cycle. Figure 2 shows the conflict graph of these 12 transformations, where normal lines indicate conflicting pairs, and dotted lines indicate conflicting triples. To save space we use three digits to represent each transformation, for example, 112 stands for the transformation $[1, 1, 2]$, and $(112)(113) = 111$. We can choose at most two inputs from each triple and at most one from each conflicting pair. So there are at most 6 conflict-free transformations from the 12, for example, b_1, b_3, b_4, b_5, $c_1 = [1, 1, 2]$, $c_2 = [2, 3, 3]$. Adding a_0, a_1, a_2 and a_3, we get a total of at most 10. The inputs a_0, b_4, b_5, c_1 are conflict-free and generate precisely these 10 transformations. Hence $\sigma(L) \leqslant 10$ for any star-free language L with $\kappa(L) = n = 3$, and this bound is tight.

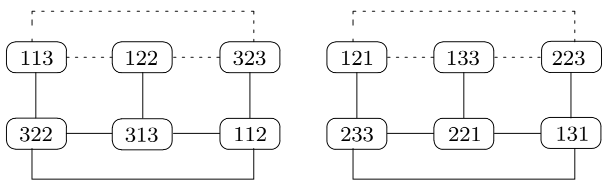

Fig. 2. Conflict graph for $n = 3$

4 Monotonicity in Transformations, Automata and Languages

We now study syntactic semigroups of languages accepted by monotonic and related automata. We denote by C_k^n the binomial coefficient "n choose k".

4.1 Monotonic Transformations, DFA's and Languages

We have shown that the tight upper bound for $n = 3$ is 10, and it turns out that this bound is met by a monotonic language (defined below). This provides one reason to study monotonic automata and languages. A second reason is the fact that all the tight upper bounds on the quotient complexity of operations on star-free languages are met by monotonic languages [4].

A transformation t of Q is *monotonic* if there exists a total order $\leqslant$ on Q such that, for all $p, q \in Q$, $p \leqslant q$ implies $pt \leqslant qt$. From now on we assume that $\leqslant$ is the usual order on integers, and that $p < q$ means that $p \leqslant q$ and $p \neq q$.

Let M_Q be the set of all monotonic transformations of Q. In the following, we restate slightly the result of Gomes and Howie [8,11] for our purposes, since the work in [8] does not consider the identity transformation to be monotonic.

Theorem 3 (Gomes and Howie). *When $n \geqslant 1$, the set M_Q is an aperiodic semigroup of cardinality*

$$|M_Q| = f(n) = \sum_{k=1}^{n} C_{k-1}^{n-1} C_k^n = C_n^{2n-1},$$

and it is generated by the set $H = \{a, b_1, \ldots, b_{n-1}, c\}$, where, for $1 \leqslant i \leqslant n-1$,
1. *$1a = 1$, $ja = j - 1$ for $2 \leqslant j \leqslant n$;*
2. *$ib_i = i + 1$, and $jb_i = j$ for all $j \neq i$;*
3. *c is the identity transformation.*

Moreover, for $n = 1$, a and c coincide and the cardinality of the generating set cannot be reduced for $n \geqslant 2$.

Remark 4. By Stirling's approximation, $f(n) = |M_Q|$ grows asymptotically like $4^n/\sqrt{\pi n}$ as $n \to \infty$.

Now we turn to DFA's whose inputs perform monotonic transformations. A DFA is *monotonic* [1] if all transformations in its transition semigroup are monotonic with respect to some fixed total order. Every monotonic DFA is aperiodic because monotonic transformations are aperiodic. A regular language is *monotonic* if its quotient DFA is monotonic.

Let us now define a DFA having as inputs the generators of M_Q:

Definition 5. *For $n \geqslant 1$, let $\mathcal{A}_n = (Q, \Sigma, \delta, 1, \{1\})$ be the DFA in which $Q = \{1, \ldots, n\}$, $\Sigma = \{a, b_1, \ldots, b_{n-1}, c\}$, and each letter in Σ performs the transformation defined in Theorem 3.*

DFA $\mathcal{A}_n$ is minimal, since state 1 is the only accepting state, and for $2 \leqslant i \leqslant n$ only state i accepts a^{i-1}. From Theorem 3 we have

Corollary 6. *For $n \geqslant 1$, the syntactic complexity $\sigma(L)$ of any monotonic language L with n quotients satisfies $\sigma(L) \leqslant f(n) = C_n^{2n-1}$. Moreover, this bound is met by the language $L(\mathcal{A}_n)$ of Definition 5, and, when $n \geqslant 2$, it cannot be met by any monotonic language over an alphabet having fewer than $n + 1$ letters.*

4.2 Monotonic Partial Transformations and IDFA's

As we shall see, for $n \geqslant 4$ the maximal syntactic complexity cannot be reached by monotonic languages; hence we continue our search for larger semigroups of aperiodic transformations. In this subsection, we extend the concept of monotonicity from full transformations to partial transformations, and hence define a new subclass of star-free languages. The upper bound of syntactic complexity of languages in this subclass is above that of monotonic languages for $n \geqslant 4$.

A partial transformation t of Q is *monotonic* if there exists a total order $\leqslant$ on Q such that, for all $p, q \in \mathrm{dom}(t)$, $p \leqslant q$ implies $pt \leqslant qt$. As before, we assume that the total order on Q is the usual order on integers. Let PM_Q be the set of all monotonic partial transformations of Q with respect to such an order. Gomes and Howie [8] showed the following result, again restated slightly:

Theorem 7 (Gomes and Howie). *When $n \geqslant 1$, the set PM_Q is an aperiodic semigroup of cardinality*

$$|PM_Q| = g(n) = \sum_{k=0}^{n} C_k^n C_k^{n+k-1},$$

and it is generated by the set $I = \{a, b_1, \dots, b_{n-1}, c_1, \dots, c_{n-1}, d\}$, where, for $1 \leqslant i \leqslant n-1$,

1. *$1a = \Box$, and $ja = j - 1$ for $j = 2, \dots, n$;*
2. *$ib_i = i + 1$, $(i+1)b_i = \Box$, and $jb_i = j$ for $j = 1, \dots, i-1, i+2, \dots, n$;*
3. *$ic_i = i + 1$, and $jc_i = j$ for all $j \neq i$;*
4. *d is the identity transformation.*

Moreover, the cardinality of the generating set cannot be reduced.

Example 8. For $n = 1$, the two monotonic partial transformations are $a = [\Box]$, and $d = [1]$. For $n = 2$, the eight monotonic partial transformations are generated by $a = [\Box, 1]$, $b_1 = [2, \Box]$, $c_1 = [2, 2]$, and $d = [1, 2]$. For $n = 3$, the 38 monotonic partial transformations are generated by $a = [\Box, 1, 2]$, $b_1 = [2, \Box, 3]$, $b_2 = [1, 3, \Box]$, $c_1 = [2, 2, 3]$, $c_2 = [1, 3, 3]$ and $d = [1, 2, 3]$.

Partial transformations correspond to IDFA's. For example, $a = [\Box, 1]$, $b = [2, \Box]$ and $c = [2, 2]$ correspond to the transitions of the IDFA of Fig. 3 (a). ■

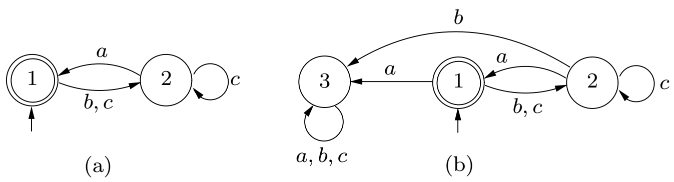

Fig. 3. Partially monotonic automata: (a) IDFA; (b) DFA

Laradji and Umar [13] proved the following asymptotic approximation:

Remark 9. For large n, $g(n) = |PM_Q| \sim 2^{-3/4}(\sqrt{2} + 1)^{2n+1}/\sqrt{\pi n}$.

An IDFA is *monotonic* if all partial transformations in its transition semigroup are monotonic with respect to some fixed total order. A quotient DFA is *partially monotonic* if its corresponding quotient IDFA is monotonic. A regular language is *partially monotonic* if its quotient DFA is partially monotonic. Note that monotonic languages are also partially monotonic.

Example 10. If we complete the transformations in Fig. 3 (a) by replacing the undefined entry □ by a new empty (or "sink") state 3, as usual, we obtain the DFA of Fig. 3 (b). That DFA is not monotonic, because $1 < 2$ implies $2 < 3$ under input b and $3 < 2$ under ab. A contradiction is also obtained if we assume that $2 < 1$. However, this DFA is partially monotonic, since its corresponding IDFA, shown in Fig. 3 (a), is monotonic.

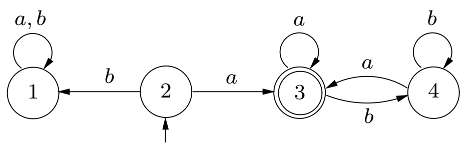

Fig. 4. Partially monotonic DFA that is monotonic and has an empty state

The DFA of Fig. 4 is monotonic for the order shown. It has an empty state, and is also partially monotonic for the same order. ■

Consider any partially monotonic language L with quotient complexity n. If its quotient DFA $\mathcal{A}$ does not have the empty quotient, then L is monotonic; otherwise, its quotient IDFA $\mathcal{I}$ has $n-1$ states, and the transition semigroup of $\mathcal{I}$ is a subset of $PM_{Q'}$, where $Q' = \{1, \dots, n-1\}$. Hence we consider the following semigroup CM_Q of *monotonic completed transformations* of Q. Start with the semigroup $PM_{Q'}$. Convert all $t \in PM_{Q'}$ to full transformations by adding n to $\mathrm{dom}(t)$ and letting $it = n$ for all $i \in Q \setminus \mathrm{dom}(t)$. Such a conversion provides a one-to-one correspondence between $PM_{Q'}$ and CM_Q. For $n \geqslant 2$, let $e(n) = g(n-1)$. Then semigroups CM_Q and $PM_{Q'}$ are isomorphic, and $e(n) = |CM_Q|$.

Definition 11. *For $n \geqslant 1$, let $\mathcal{B}_n = (Q, \Sigma, \delta, 1, \{1\})$ be the DFA in which $Q = \{1, \dots, n\}$, $\Sigma = \{a, b_1, \dots, b_{n-2}, c_1, \dots, c_{n-2}, d\}$, and each letter in Σ defines a transformation such that, for $1 \leqslant i \leqslant n-2$,*

1. *$1a = na = n$, and $ja = j-1$ for $j = 2, \dots, n-1$;*
2. *$ib_i = i+1$, $(i+1)b_i = n$, and $jb_i = j$ for $j = 1, \dots, i-1, i+2, \dots, n$;*
3. *$ic_i = i+1$, and $jc_i = j$ for all $j \neq i$;*
4. *d is the identity transformation.*

We know that monotonic languages are also partially monotonic. As shown in Table 1, $|M_Q| = f(n) > e(n) = |CM_Q|$ for $n \leqslant 3$. On the other hand, one verifies that $e(n) > f(n)$ when $n \geqslant 4$. By Corollary 6 and Theorem 7, we have

Corollary 12. *The syntactic complexity of a partially monotonic language L with n quotients satisfies $\sigma(L) \leqslant f(n)$ for $n \leqslant 3$, and $\sigma(L) \leqslant e(n)$ for $n \geqslant 4$. Moreover, when $n \geqslant 4$, this bound is met by $L(\mathcal{B}_n)$ of Definition 11, and it cannot be met by any partially monotonic language over an alphabet having fewer than $2n-2$ letters.*

Table 1 contains these upper bounds for small values of n. By Remark 9, the upper bound $e(n)$ is asymptotically $2^{-3/4}(\sqrt{2}+1)^{2n-1}/\sqrt{\pi(n-1)}$.

4.3 Nearly Monotonic Transformations and DFA's

In this section we develop an even larger aperiodic semigroup based on partially monotonic languages.

Let K_Q be the set of all constant transformations of Q, and let $NM_Q = CM_Q \cup K_Q$. We shall call the transformations in NM_Q *nearly monotonic* with respect to the usual order on integers.

Theorem 13. *When $n \geqslant 2$, the set NM_Q of all nearly monotonic transformations of a set Q of n elements is an aperiodic semigroup of cardinality*

$$|NM_Q| = h(n) = e(n) + (n-1) = \sum_{k=0}^{n-1} C_k^{n-1} C_k^{n+k-2} + (n-1),$$

and it is generated by the set $J = \{a, b_1, \ldots, b_{n-2}, c_1, \ldots, c_{n-2}, d, e\}$ of $2n-1$ transformations of Q, where e is the constant transformation $\binom{Q}{1}$, and all other transformations are as in Definition 11. Moreover, the cardinality of the generating set cannot be reduced.

Proof. Pick any $t_1, t_2 \in NM_Q$. If $t_1, t_2 \in CM_Q$, then $t_1t_2, t_2t_1 \in CM_Q$. Otherwise $t_1 \in K_Q$ or $t_2 \in K_Q$, and both t_1t_2, t_2t_1 are constant transformations. Hence $t_1t_2, t_2t_1 \in NM_Q$ and NM_Q is a semigroup. Since constant transformations are aperiodic and CM_Q is aperiodic, NM_Q is also aperiodic.

If X is a set of transformations, let $\langle X \rangle$ denote the semigroup generated by X. Since $J \subseteq NM_Q$, $\langle J \rangle \subseteq NM_Q$. Let $I' = J \setminus \{e\}$, and $Q' = Q \setminus \{n\}$. Then $PM_{Q'} \simeq CM_Q = \langle I' \rangle$. For any $t = \binom{Q}{j} \in K_Q$, where $j \in Q$, since $s_j = \binom{Q}{j}\binom{n}{n} \in CM_Q \subseteq \langle J \rangle$, we have that $t = es_j \in \langle J \rangle$. So $NM_Q = \langle J \rangle$. Note that $\binom{Q}{i} \in CM_Q$ if and only if $i = n$. Thus $h(n) = |NM_Q| = |PM_{Q'}| + (n-1) = e(n) + (n-1)$.

Since the cardinality of I' cannot be reduced, and $e \notin \langle I' \rangle$, also the cardinality of J cannot be reduced. □

An input $a \in \Sigma$ is *constant* if it performs a constant transformation of Q. Let $\mathcal{A}$ be a DFA with alphabet Σ; then $\mathcal{A}$ is *nearly monotonic* if, after removing constant inputs, the resulting DFA $\mathcal{A}'$ is partially monotonic. A regular language is *nearly monotonic* if its quotient DFA is nearly monotonic.

Definition 14. *For $n \geqslant 2$, let $\mathcal{C}_n = (Q, \Sigma, \delta, 1, \{1\})$ be a DFA, where $Q = \{1, \ldots, n\}$, $\Sigma = \{a, b_1, \ldots, b_{n-2}, c_1, \ldots, c_{n-2}, d, e\}$, and each letter in Σ performs the transformation defined in Theorem 13 and Definition 11.*

Theorem 13 now leads us to the following result:

Theorem 15. *For $n \geqslant 2$, if L is a nearly monotonic language L with n quotients, then $\sigma(L) \leqslant h(n) = \sum_{k=0}^{n-1} C_k^{n-1} C_k^{n+k-2} + (n-1)$. Moreover, this bound is met by the language $L(\mathcal{C}_n)$ of Definition 14, and cannot be met by any nearly monotonic language over an alphabet having fewer than $2n-1$ letters.*

Proof. State 1 is reached by ε. For $2 \leqslant i \leqslant n-1$, state i is reached by $w_i = b_1 \cdots b_{i-1}$. State n is reached by $w_{n-1}b_{n-2}$. Thus all states are reachable. For $1 \leqslant i \leqslant n-1$, the word a^{i-1} is only accepted by state i. Also, state n rejects a^i for all $i \geqslant 0$. So all n states are distinguishable, and $\mathcal{C}_n$ is minimal. Thus L has n quotients. The syntactic semigroup of L is generated by J; so L has syntactic complexity $\sigma(L) = h(n) = \sum_{k=0}^{n-1} C_k^{n-1} C_k^{n+k-2} + (n-1)$, and it is star-free. □

As shown earlier, $e(n) > f(n)$ for $n \geqslant 4$. Since $h(n) = e(n) + (n-1)$, and $h(n) = f(n)$ for $n \in \{2, 3\}$, as shown in Table 1, we have that $h(n) \geqslant f(n)$ for $n \geqslant 2$, and the maximal syntactic complexity of nearly monotonic languages is at least that of both monotonic and partially monotonic languages.

Although we cannot prove that NM_Q is the largest semigroup of aperiodic transformations, we can show that no transformation can be added to NM_Q.

A set $\mathbb{S} = \{T_1, T_2, \ldots, T_k\}$ of transformation semigroups is a *chain* if $T_1 \subset T_2 \subset \cdots \subset T_k$. Semigroup T_k is the largest in $\mathbb{S}$, and we denote it by $\max(\mathbb{S}) = T_k$. The following result shows that the syntactic semigroup $S_{L(\mathcal{C}_n)} = T_{\mathcal{C}_n}$ of $L(\mathcal{C}_n)$ in Definition 14 is a local maximum among aperiodic subsemigroups of $\mathcal{T}_Q$.

Proposition 16. *Let $\mathbb{S}$ be a chain of aperiodic subsemigroups of $\mathcal{T}_Q$. If $T_{\mathcal{C}_n} \in \mathbb{S}$, then $T_{\mathcal{C}_n} = \max(\mathbb{S})$.*

Proof. Suppose $\max(\mathbb{S}) = T_k$ for some aperiodic subsemigroup T_k of $\mathcal{T}_Q$, and $T_k \neq T_{\mathcal{C}_n}$. Then there exist $t \in T_k$ such that $t \notin T_{\mathcal{C}_n}$, and $i, j \in Q$ such that $i < j \neq n$ but $it > jt$, and $it, jt \neq n$. Let $\tau \in \mathcal{T}_Q$ be such that $(jt)\tau = i$, $(it)\tau = j$, and $h\tau = n$ for all $h \neq i, j$; then $\tau \in T_{\mathcal{C}_n}$. Let $\lambda \in \mathcal{T}_Q$ be such that $i\lambda = i$, $j\lambda = j$, and $h\lambda = n$ for all $h \neq i, j$; then also $\lambda \in T_{\mathcal{C}_n}$. Since $T_k = \max(\mathbb{S})$, $T_{\mathcal{C}_n} \subset T_k$ and $\tau, \lambda \in T_k$. Then $s = \lambda t \tau$ is also in T_k. However, $is = i(\lambda t \tau) = j$, $js = j(\lambda t \tau) = i$, and $hs = n$ for all $h \neq i, j$; then $s = (i, j)\binom{P}{n}$, where $P = Q \setminus \{i, j\}$, is not aperiodic, a contradiction. Therefore $T_{\mathcal{C}_n} = \max(\mathbb{S})$. □

5 Conclusions

We conjecture that the syntactic complexity of languages accepted by the nearly monotonic DFA's of Definition 14 meets the upper bound for star-free languages:

Conjecture 17. The syntactic complexity of a star-free language L with $\kappa(L) = n \geqslant 4$ satisfies $\sigma(L) \leqslant h(n)$.

Our results are summarized in Table 1. Let $Q = \{1, \ldots, n\}$, and $Q' = Q \setminus \{n\}$. The figures in bold type are tight bounds verified using *GAP* [7], by enumerating aperiodic subsemigroups of $\mathcal{T}_Q$. The asterisk $*$ indicates that the bound is already tight for a smaller alphabet. The last five rows show the values of $f(n) = |M_Q|$, $e(n) = |CM_Q| = g(n-1) = |PM_{Q'}|$, $h(n) = |NM_Q|$, and the weak upper bound $(n+1)^{n-1}$.

Table 1. Syntactic complexity of star-free languages

$\|\Sigma\|$ / n	1	2	3	4	5	6
1	**1**	**1**	**2**	**3**	**5**	**6**
2	$*$	**2**	**7**	**19**	**62**	?
3	$*$	**3**	**9**	**31**	?	?
4	$*$	$*$	**10**	**34**	?	?
5	$*$	$*$	$*$	37	125	?
...	...	...	...	...	...	...
$f(n) = \|M_Q\|$	**1**	**3**	**10**	35	126	462
$e(n) = \|CM_Q\| = g(n-1) = \|PM_{Q'}\|$	–	2	8	38	192	1,002
$h(n) = \|NM_Q\| = e(n) + (n-1)$	–	**3**	**10**	41	196	1,007
$(n+1)^{n-1}$	1	3	16	125	1,296	16,807

Acknowledgment. We thank Zoltan Ésik and Judit Nagy-Gyorgy for pointing out to us the relationship between aperiodic inputs and Cayley's theorem.

References

1. Ananichev, D.S., Volkov, M.V.: Synchronizing generalized monotonic automata. Theoret. Comput. Sci. 330(1), 3–13 (2005)
2. Brzozowski, J.: Quotient complexity of regular languages. J. Autom. Lang. Comb. 15(1/2), 71–89 (2010)
3. Brzozowski, J., Li, B., Ye, Y.: Syntactic complexity of prefix-, suffix-, bifix-, and factor-free regular languages. Theoret. Comput. Sci. (in press, 2012)
4. Brzozowski, J., Liu, B.: Quotient complexity of star-free languages. In: Dömösi, P., Szabolcs, I. (eds.) 13th International Conference on Automata and Formal Languages (AFL), pp. 138–152. Institute of Mathematics and Informatics, College of Nyíregyháza (2011)
5. Brzozowski, J., Ye, Y.: Syntactic Complexity of Ideal and Closed Languages. In: Mauri, G., Leporati, A. (eds.) DLT 2011. LNCS, vol. 6795, pp. 117–128. Springer, Heidelberg (2011)
6. Caley, A.: A theorem on trees. Quart. J. Math. 23, 376–378 (1889)
7. GAP-Group: GAP - Groups, Algorithms, Programming - a System for Computational Discrete Algebra (2010), `http://www.gap-system.org/`
8. Gomes, G., Howie, J.: On the ranks of certain semigroups of order-preserving transformations. Semigroup Forum 45, 272–282 (1992)
9. Holzer, M., König, B.: On deterministic finite automata and syntactic monoid size. Theoret. Comput. Sci. 327(3), 319–347 (2004)
10. Holzer, M., Kutrib, M., Meckel, K.: Nondeterministic state complexity of star-free languages. Theoret. Comput. Sci. (in press, 2012)
11. Howie, J.M.: Products of idempotents in certain semigroups of transformations. Proc. Edinburgh Math. Soc. 17(2), 223–236 (1971)
12. Krawetz, B., Lawrence, J., Shallit, J.: State complexity and the monoid of transformations of a finite set (2003), `http://arxiv.org/abs/math/0306416v1`
13. Laradji, A., Umar, A.: Asymptotic results for semigroups of order-preserving partial transformations. Comm. Alg. 34(3), 1071–1075 (2006)

14. Maslov, A.N.: Estimates of the number of states of finite automata. Dokl. Akad. Nauk SSSR 194, 1266–1268 (1970) (Russian); english translation: Soviet Math. Dokl. 11, 1373–1375 (1970)
15. McNaughton, R., Papert, S.A.: Counter-Free Automata. M.I.T. research monograph, vol. 65. The MIT Press (1971)
16. Myhill, J.: Finite automata and the representation of events. Wright Air Development Center Technical Report 57–624 (1957)
17. Nerode, A.: Linear automaton transformations. Proc. Amer. Math. Soc. 9, 541–544 (1958)
18. Rozenberg, G., Salomaa, A. (eds.): Handbook of Formal Languages, vol. 1: Word, Language, Grammar. Springer, New York (1997)
19. Schützenberger, M.P.: On finite monoids having only trivial subgroups. Inform. and Control 8, 190–194 (1965)
20. Shor, P.W.: A new proof of Cayley's formula for counting labeled trees. J. Combin. Theory Ser. A 71(1), 154–158 (1995)
21. Yu, S.: State complexity of regular languages. J. Autom. Lang. Comb. 6, 221–234 (2001)