

# Decision Problems For Convex Languages

Janusz Brzozowski, Jeffrey Shallit, and Zhi Xu

David R. Cheriton School of Computer Science,  
University of Waterloo, Waterloo, ON, Canada N2L 3G1  
{brzozo, shallit, z5xu}@uwaterloo.ca

**Abstract.** We examine decision problems for various classes of convex languages, previously studied by Ang and Brzozowski, originally under the name “continuous languages”. We can decide whether a language  $L$  is prefix-, suffix-, factor-, or subword-convex in polynomial time if  $L$  is represented by a DFA, but these problems become PSPACE-complete if  $L$  is represented by an NFA. If a regular language is not convex, we find tight upper bounds on the length of the shortest words demonstrating this fact, in terms of the number of states of an accepting DFA. Similar results are proved for some subclasses of convex languages: the prefix-, suffix-, factor-, and subword-closed languages, and the prefix-, suffix-, factor-, and subword-free languages. Finally, we briefly examine these questions where  $L$  is represented by a context-free grammar.

## 1 Introduction

A word  $x$  is a *factor* of a word  $w$  if  $w = uxv$  for some words  $u$  and  $v$ . If in addition  $u = \varepsilon$ , the empty word, then  $x$  is a *prefix* of  $w$ ; if  $v = \varepsilon$ , then  $x$  is a *suffix* of  $w$ . A word  $x$  is a *subword* of  $w$  if  $x$  can be obtained by striking out zero or more letters of  $w$ , that is, if there exist words  $w_0, w_1, \dots, w_n, x_1, x_2, \dots, x_n$  such that  $w = w_0x_1w_1x_2 \cdots x_nw_n$  and  $x = x_1x_2 \cdots x_n$ . (In the literature, what we call “factor” and “subword” are sometimes called “subword” and “subsequence”, respectively, which can create confusion. We emphasize that in this paper, a factor is a contiguous block, while a subword can be “scattered”.) A factor  $x$  (respectively, prefix, suffix, subword) of  $w$  is *proper* if  $x \neq w$ .

We consider some computational complexity questions concerning several classes of convex languages. The concept of a language convex with respect to the subword relation was introduced by Thierrin [1] in 1973. In our terminology, a language  $L$  is *subword-convex* if  $u, w \in L$  with  $u$  a subword of  $w$  implies that any word  $v$  must also be in  $L$  if  $u$  is a subword of  $v$  and  $v$  is a subword of  $w$ . A language  $L$  is *subword-free* if  $w \in L$  implies that no proper subword of  $w$  is in  $L$ . A language  $L$  is *subword-closed* if  $w \in L$  implies that every subword of  $w$  is also in  $L$ . Finally, a language is *converse-subword-closed* if  $w \in L$  implies that every word  $v$  that has  $w$  as a subword is also in  $L$ . Subword-free, subword-closed and converse-subword-closed languages are special cases of subword-convex languages.

Recently Ang and Brzozowski [2, 3] generalized the concepts of convex, free, and closed languages to arbitrary relations. Of particular interest are the prefix,

suffix, factor, and subword relations. The definitions of prefix-, suffix-, and factor-convex languages, prefix-, suffix-, and factor-free languages, and prefix-, suffix-, and factor-closed languages are similar to those for the subword relation. Besides fitting in a common mathematical framework, convex languages are of interest because they include several well-studied classes of languages. Thus, prefix-free (respectively, suffix-free, factor-free, and subword-free) languages are prefix codes (respectively, suffix codes, infix codes, and hypercodes) [4, 5]. Factor-closed languages, often called *factorial*, were studied by de Luca and Varricchio [6], who showed that every factor-closed language is a complement of a two-sided ideal, that is, of a language  $L$  over an alphabet  $\Sigma$ , satisfying  $L = \Sigma^*L\Sigma^*$ . It was noted in [2, 3] that prefix-closed languages are complements of right ideals and suffix-closed languages are complements of left ideals. Moreover, every subword-closed language is a complement of a language  $L$  which is equal to the shuffle of  $L$  with  $\Sigma^*$ .

In this paper, we consider the computational complexity of testing whether a given language is prefix-convex, suffix-convex, etc., prefix-closed, suffix-closed, etc., prefix-free, suffix-free, etc., for a total of 12 different problems. As we will see, the computational complexity of these decision problems depends on how the language is represented. If it is specified by a DFA (deterministic finite automaton), each decision problem is solvable in polynomial time. If the language is represented as a regular expression or an NFA (nondeterministic finite automaton), the decision problems for closure and convexity are PSPACE-complete, but still solvable in polynomial time for freeness. We also consider the following question: given that a language is *not* prefix-convex, suffix-convex, etc., what is a good upper bound on the length of the shortest words (*witnesses*) demonstrating this fact?

The remainder of the paper is structured as follows. In Section 2 we study the complexity of testing for convexity for languages represented by DFA's, and include testing for closure and freeness as special cases. In Section 3 we exhibit shortest witnesses to the lack of convexity. In Section 4 we prove that our decision problems for DFA's are NL-complete. Convex languages specified by NFA's and context-free grammars are briefly studied in Section 5. Section 6 concludes the paper.

This is an extended version of our paper from the LATA 2009 conference [7, 8].

## 2 Decision problems for languages specified by DFA's

We will show that, if a regular language  $L$  is represented by a DFA  $M$  with  $n$  states, it is possible to test prefix-, suffix-, factor-, and subword-convexity efficiently, in fact, in  $O(n^3)$  time.

Let  $\triangleleft$  be one of the four relations *prefix*, *suffix*, *factor*, or *subword*. The basic idea is as follows:  $L$  is *not*  $\triangleleft$ -convex if and only if there exist words  $u, w \in L$ ,  $v \notin L$ , such that  $u \triangleleft v \triangleleft w$ . Given  $M$ , we create an NFA- $\varepsilon$  (an NFA allowing transitions on the empty word  $\varepsilon$ )  $M'$  with  $O(n^3)$  states and transitions that

accepts the language

$$\{w \in L(M) : \text{there exist } u \in L(M), v \notin L(M) \text{ such that } u \preceq v \preceq w\}.$$

Then  $L(M') = \emptyset$  if and only if  $L(M)$  is  $\preceq$ -convex. We can test the emptiness of  $L(M')$  using depth-first search in time linear in the size of  $M'$ . This gives an  $O(n^3)$  algorithm for testing  $\preceq$ -convexity.

Since the constructions for all four properties are similar, we handle the hardest case (factor-convexity) in detail. We content ourselves with a brief sketch of the necessary constructions.

## 2.1 Factor-convex languages

Suppose  $M = (Q, \Sigma, \delta, q_0, F)$  is a DFA accepting the language  $L = L(M)$ , and suppose  $M$  has  $n$  states. We construct an NFA- $\varepsilon$   $M'$  with the property that  $L(M')$  is the set of words  $w \in \Sigma^*$  such that there exist  $u, v \in \Sigma^*$  where  $u$  is a factor of  $v$ ,  $v$  is a factor of  $w$ , and  $u, w \in L, v \notin L$ . Clearly  $L(M') = \emptyset$  if and only if  $L(M)$  is factor-convex.

Here is the construction of  $M'$ . States of  $M'$  are quadruples, where components 1, 2, and 3 keep track of the state of  $M$  as it is processing  $w$ ,  $v$ , and  $u$  (respectively). The last component is a flag indicating the present *mode* of the simulation process. Formally,  $M' = (Q', \Sigma, \delta', q'_0, F')$ , where  $Q' = Q \times Q \times Q \times \{1, 2, 3, 4, 5\}$ ,  $q'_0 = [q_0, q_0, q_0, 1]$ ,  $F' = F \times (Q - F) \times F \times \{5\}$ , and

1.  $\delta'([p, q_0, q_0, 1], a) = \{[\delta(p, a), q_0, q_0, 1]\}$ , for all  $p \in Q, a \in \Sigma$ ;
2.  $\delta'([p, q_0, q_0, 1], \varepsilon) = \{[p, q_0, q_0, 2]\}$ , for all  $p \in Q$ ;
3.  $\delta'([p, q, q_0, 2], a) = \{[\delta(p, a), \delta(q, a), q_0, 2]\}$ , for all  $p, q \in Q, a \in \Sigma$ ;
4.  $\delta'([p, q, q_0, 2], \varepsilon) = \{[p, q, q_0, 3]\}$ , for all  $p, q \in Q$ ;
5.  $\delta'([p, q, r, 3], a) = \{[\delta(p, a), \delta(q, a), \delta(r, a), 3]\}$ , for all  $p, q, r \in Q, a \in \Sigma$ ;
6.  $\delta'([p, q, r, 3], \varepsilon) = \{[p, q, r, 4]\}$ , for all  $p, q, r \in Q$ ;
7.  $\delta'([p, q, r, 4], a) = \{[\delta(p, a), \delta(q, a), r, 4]\}$ , for all  $p, q, r \in Q, a \in \Sigma$ ;
8.  $\delta'([p, q, r, 4], \varepsilon) = \{[p, q, r, 5]\}$ , for all  $p, q, r \in Q$ ;
9.  $\delta'([p, q, r, 5], a) = \{[\delta(p, a), q, r, 5]\}$ , for all  $p, q, r \in Q, a \in \Sigma$ .

One can verify that the NFA- $\varepsilon$   $M'$  has  $3n^3 + n^2 + n$  states and  $(3|\Sigma| + 2)n^3 + (|\Sigma| + 1)(n^2 + n)$  transitions, where  $|\Sigma|$  is the cardinality of  $\Sigma$ .

To see that the construction is correct, suppose  $L$  is not factor-convex. Then there exist words  $u, v, w$  such that  $u$  is a factor of  $v$ ,  $v$  is a factor of  $w$ , and  $u, w \in L$  while  $v \notin L$ . Then there exist words  $u', u'', v', v''$  such that  $v = u'uu''$  and  $w = v'vv'' = v'u'uu''v''$ . Let  $\delta(q_0, v') = q_1$ ,  $\delta(q_1, u') = q_2$ ,  $\delta(q_2, u) = q_3$ ,  $\delta(q_3, u'') = q_4$ , and  $\delta(q_4, v'') = q_5$ . Moreover, let  $\delta(q_0, u') = q_a$ ,  $\delta(q_a, u) = q_b$ , and  $\delta(q_b, u'') = q_c$ , and  $\delta(q_0, u) = q_\alpha$ . Since  $u, w \in L$ , we know that  $q_\alpha$  and  $q_5$  are accepting states. Since  $v \notin L$ , we know that  $q_c$  is not accepting.

The automaton  $M'$  operates as follows. In the initial state  $[q_0, q_0, q_0, 1]$  we process the symbols of  $v'$  using Rule 1, ending in the state  $[q_1, q_0, q_0, 1]$ . At this

point, we use Rule 2 to move to  $[q_1, q_0, q_0, 2]$  by an  $\varepsilon$ -move. Next, we process the symbols of  $u'$  using Rule 3, ending in the state  $[q_2, q_a, q_0, 2]$ . Then we use Rule 4 to move to  $[q_2, q_a, q_0, 3]$  by an  $\varepsilon$ -move. Next, we process the symbols of  $u$  using Rule 5, ending in the state  $[q_3, q_b, q_a, 3]$ . Then we use Rule 6 to move to  $[q_3, q_b, q_a, 4]$  by an  $\varepsilon$ -move. Next, we process the symbols of  $u''$  using Rule 7, ending in the state  $[q_4, q_c, q_a, 4]$ . Then we use Rule 8 to move to  $[q_4, q_c, q_a, 5]$  by an  $\varepsilon$ -move. Finally, we process the symbols of  $v''$  using Rule 9, ending in the state  $[q_5, q_c, q_a, 5]$ , and this state is in  $F'$ .

On the other hand, suppose  $M'$  accepts the input  $w$ . Then we must have  $\delta'(q'_0, w) \cap F' \neq \emptyset$ . But, by our construction, the only way to reach a state in  $F'$  is to apply Rules 1 through 9 in that order, where odd-numbered rules can be used any number of times, and even-numbered rules can be used only once. Letting  $v', u', u, u'', v''$  be the words labeling the uses of Rules 1, 3, 5, 7, and 9, respectively, we see that  $w = v'u'uu''v''$ , where  $\delta(q_0, w) \in L$ ,  $\delta(q_0, u) \in L$ , and  $\delta(q_0, u'uu'') \notin L$ . It follows that  $u, w \in L$  and  $v = u'uu'' \notin L$ , and so  $L$  is not factor-convex.

We have proved the following theorem:

**Theorem 1.** *If  $M$  is a DFA with  $n$  states, there exists an NFA- $\varepsilon$   $M'$  with  $O(n^3)$  states and transitions such that  $M'$  accepts the language*

$$L(M') = \{w \in \Sigma^* : w \in L \text{ and there exist } u, v \in \Sigma^* \text{ such that } u \text{ is a factor of } v, \\ v \text{ is a factor of } w, \text{ and } u \in L, v \notin L\}.$$

**Corollary 1.** *We can decide if a given regular language  $L$  accepted by a DFA with  $n$  states is factor-convex in  $O(n^3)$  time.*

*Proof.* Since  $L$  is factor-convex if and only if  $L(M') = \emptyset$ , it suffices to check if  $L(M') = \emptyset$  using depth-first search of a directed graph, in time linear in the number of vertices and edges of  $M'$ .  $\square$

**Factor-closed languages** The language  $L$  is *not* factor-closed if and only if there exist words  $v, w$  such that  $v$  is a factor of  $w$ , and  $w \in L$ , while  $v \notin L$ . Given a DFA  $M$  accepting  $L$ , we construct an NFA- $\varepsilon$   $M'$  such that

$$L(M') = \{w \in \Sigma^* : w \in L \text{ and there exists } v \in \Sigma^* \text{ such that } v \text{ is a factor of } w, \\ \text{and } v \notin L\}.$$

Then  $L(M') = \emptyset$  if and only if  $L(M)$  is factor-closed.

States of  $M'$  are triples, where components 1 and 2 keep track of the state of  $M$  as it is processing  $w$  and  $v$  (respectively). The last component is a flag as before. Formally,  $M' = (Q', \Sigma, \delta', q'_0, F')$ , where  $Q' = Q \times Q \times \{1, 2, 3\}$ ;  $q'_0 = [q_0, q_0, 1]$ ;  $F' = F \times (Q - F) \times \{3\}$ ; and

1.  $\delta'([p, q_0, 1], a) = \{\delta(p, a), q_0, 1\}$  for  $p \in Q$ ,  $a \in \Sigma$ .
2.  $\delta'([p, q_0, 1], \varepsilon) = \{[p, q_0, 2]\}$ , for all  $p \in Q$ ;

3.  $\delta'([p, q, 2], a) = \{[\delta(p, a), \delta(q, a), 2]\}$ , for all  $p, q \in Q$ ;
4.  $\delta'([p, q, 2], \varepsilon) = \{[p, q, 3]\}$ , for all  $p, q \in Q$ ;
5.  $\delta'([p, q, 3], a) = \{[\delta(p, a), q, 3]\}$ , for  $p, q \in Q, a \in \Sigma$ .

$M'$  has  $2n^2 + n$  states and  $(2|\Sigma| + 1)n^2 + (|\Sigma| + 1)$  transitions. Thus we have

**Theorem 2.** *We can decide if a given regular language  $L$  accepted by a DFA with  $n$  states is factor-closed in  $O(n^2)$  time.*

This result was previously obtained by Béal et al. [9, Prop. 5.1, p. 13] through a slightly different approach.

The converse of the relation “ $u$  is a factor of  $v$ ” is “ $v$  contains  $u$  as a factor”. This relation and similar converse relations derived from the prefix, suffix, and subword relations, lead to “converse-closed languages” [2]. Subword-closed and converse-subword-closed languages were characterized by Thierrin [1]. It has been shown by de Luca and Varricchio [6] that a language  $L$  is factor-closed (factorial, in their terminology) if and only if it is a complement of an ideal, that is, if and only if  $L = \overline{\Sigma^* K \Sigma^*}$  for some  $K \subseteq \Sigma^*$ . Ang and Brzozowski [2] noted that a language is an ideal if and only if it is converse-factor-closed, that is, if, for every  $u \in L$ , each word of the form  $v = xuy$  is also in  $L$ . Thus, to test whether  $L$  is converse-factor-closed, we must check that there is no pair  $(u, v)$  such that  $u \in L, v \notin L$ , and  $u$  is a factor of  $v$ . This is equivalent to testing whether  $\overline{L}$  is factor-closed. Then the following is an immediate consequence of Theorem 1:

**Corollary 2.** *We can decide if a given regular language  $L$  accepted by a DFA with  $n$  states is an ideal in  $O(n^2)$  time.*

The results above apply to other converse-closed languages.

**Factor-free languages** Factor-free (also known as infix-free) languages have been studied recently by Han et al. [10], who gave efficient algorithms for determining if the language accepted by an NFA is prefix-, suffix-, or factor-free.

*Remark 1.* We can decide whether a DFA language is factor-free in  $O(n^2)$  time with the automaton we used for testing factor-closure, except that the set of accepting states is now  $F' = F \times F \times \{3\}$ .

*Remark 2.* Similar results hold for prefix-free, suffix-free, and subword-free languages.

## 2.2 Prefix-convex languages

Prefix convexity can be tested in an analogous fashion. We give the construction of  $M'$  without proof: let  $M' = (Q', \Sigma, \delta', q'_0, F')$ , where  $Q' = Q \times Q \times Q \times \{1, 2, 3\}$ ,  $q'_0 = [q_0, q_0, q_0, 1]$ ,  $F' = F \times (Q - F) \times F \times \{3\}$ , and

1.  $\delta'([p, q, r, 1], a) = \{[\delta(p, a), \delta(q, a), \delta(r, a), 1]\}$  for  $p, q, r \in Q, a \in \Sigma$ ;
2.  $\delta'([p, q, r, 1], \varepsilon) = \{[p, q, r, 2]\}$  for  $p, q, r \in Q$ ;

3.  $\delta'([p, q, r, 2], a) = \{[\delta(p, a), \delta(q, a), r, 2]\}$  for  $p, q, r \in Q, a \in \Sigma$ ;
4.  $\delta'([p, q, r, 2], \varepsilon) = \{[p, q, r, 3]\}$  for  $p, q, r \in Q$ ;
5.  $\delta'([p, q, r, 3], a) = \{[\delta(p, a), q, r, 3]\}$  for  $p, q, r \in Q, a \in \Sigma$ .

The NFA  $M'$  has  $3n^3$  states and  $3(|\Sigma| + 1)n^3$  transitions.

**Prefix-closed languages** By varying the construction in Section 2.1, we have

**Theorem 3.** *We can decide if a given regular language  $L$  accepted by a DFA with  $n$  states is prefix-closed, suffix-closed, or subword-closed in  $O(n^2)$  time.*

Previous results for prefix-, suffix-, and factor-closed languages can be found in [11, 12].

**Prefix-free languages** Already addressed in Remark 2.

### 2.3 Suffix-convex languages

Suffix-convexity can be tested in an analogous fashion. We give the construction of  $M'$  without proof. Let  $M' = (Q', \Sigma, \delta', q'_0, F')$ , where  $Q' = Q \times Q \times Q \times \{1, 2, 3\}$ ,  $q'_0 = [q_0, q_0, q_0, 1]$ ,  $F' = F \times (Q - F) \times F \times \{3\}$ , and

1.  $\delta'([p, q_0, q_0, 1], a) = \{[\delta(p, a), q_0, q_0, 1]\}$  for  $p \in Q, a \in \Sigma$ ;
2.  $\delta'([p, q_0, q_0, 1], \varepsilon) = \{[p, q_0, q_0, 2]\}$  for  $p \in Q$ ;
3.  $\delta'([p, q, q_0, 2], a) = \{[\delta(p, a), \delta(q, a), q_0, 2]\}$  for  $p, q \in Q, a \in \Sigma$ ;
4.  $\delta'([p, q, q_0, 2], \varepsilon) = \{[p, q, q_0, 3]\}$  for  $p, q \in Q$ ;
5.  $\delta'([p, q, r, 3], a) = \{[\delta(p, a), \delta(q, a), \delta(r, a), 3]\}$  for  $p, q, r \in Q, a \in \Sigma$ .

The NFA  $M'$  has  $n^3$  states and  $|\Sigma|n^3 + (|\Sigma| + 1)(n^2 + n)$  transitions.

For results on suffix-closure and suffix-freeness, see Theorem 3 and Remark 2, respectively.

### 2.4 Subword-convex languages

Subword-convexity can be tested in an analogous fashion. We give the construction of  $M'$  without proof. Let  $M' = (Q', \Sigma, \delta', q'_0, F')$ , where

$$\begin{aligned}
Q' &= Q \times Q \times Q; \\
q'_0 &= [q_0, q_0, q_0]; \\
F' &= F \times (Q - F) \times F; \\
\delta'([p, q, r], a) &= \{[\delta(p, a), q, r], [\delta(p, a), \delta(q, a), r], [\delta(p, a), \delta(q, a), \delta(r, a)]\}, \\
&\text{for all } p, q, r \in Q \text{ and } a \in \Sigma.
\end{aligned}$$

The NFA  $M'$  has  $n^3$  states and  $|\Sigma|n^3$  transitions.

The idea is that as the symbols of  $w$  are read, we keep track of the state of  $M$  in the first component. We then “guess” which symbols of the input also

belong to  $u$  and/or  $v$ , enforcing the condition that, if a symbol belongs to  $u$ , then it must belong to  $v$ , and if it belongs to  $v$ , then it must belong to  $w$ . We therefore cover all possibilities of words  $u, v$  such that  $u$  is a subword of  $v$  and  $v$  is a subword of  $w$ .

For results on subword-closure and subword-freeness, see Theorem 3 and Remark 2, respectively.

## 2.5 Almost convex languages

As we have seen, a language  $L$  is prefix-convex if and only if there are no triples  $(u, v, w)$  with  $u$  a prefix of  $v$ ,  $v$  a prefix of  $w$ , and  $u, w \in L, v \notin L$ . We call such a triple a *witness*. A language could fail to be prefix-convex because there are infinitely many witnesses (for example, the language  $(\mathbf{aa})^*$ ), or it could fail because there is at least one, but only finitely many witnesses (for example, the language  $\varepsilon + \mathbf{aaa}^*$ ).

We define a language  $L$  to be *almost prefix-convex* if there exists at least one, but only finitely many witnesses to the failure of the prefix-convex property. Analogously, we define *almost suffix-*, *almost factor-*, and *almost subword-convex*.

**Theorem 4.** *Let  $L$  be a regular language accepted by a DFA with  $n$  states. Then we can determine if  $L$  is almost prefix-convex (respectively, almost suffix-convex, almost factor-convex, almost subword-convex) in  $O(n^3)$  time.*

*Proof.* We give the proof for the almost factor-convex property, leaving the other cases to the reader.

Consider the NFA- $\varepsilon$   $M'$  defined in Section 2.1. As we have seen,  $M'$  accepts the language

$$L(M') = \{w \in \Sigma^* : \text{there exist } u, v \in \Sigma^* \text{ such that } u \text{ is a factor of } v, \\ v \text{ is a factor of } w, \text{ and } u, w \in L, v \notin L\}.$$

Then  $M'$  accepts an infinite language if and only if  $L$  is not almost factor-convex. For if  $M'$  accepts infinitely many distinct words, then there are infinitely many distinct witnesses, while if there are infinitely many distinct witnesses  $(u, v, w)$ , then there must be infinitely many distinct  $w$  among them, since the lengths of  $u$  and  $v$  are bounded by the length of  $w$ .

Thus it suffices to see if  $M'$  accepts an infinite language. If  $M'$  were an NFA, this would be trivial: first, we remove all states not reachable from the start state or from which we cannot reach a final state. Next, we look for the existence of a cycle. All three goals can be easily accomplished in time linear in the size of  $M'$ , using depth-first search.

However,  $M'$  is an NFA- $\varepsilon$ , so there is one additional complication: the cycle we find might be labeled completely by  $\varepsilon$ -transitions. To solve this, we use an idea suggested to us by Jack Zhao and Timothy Chan (personal communication): we find all the connected components of the transition graph of  $M'$  (which can be done in linear time) and then, for each edge  $(p, q)$  labeled with something

other than  $\varepsilon$  (corresponding to the transition  $q \in \delta(p, a)$  for some  $a \in \Sigma$ ), we check to see if  $p$  and  $q$  are in the same connected component. If they are, we have found a cycle labeled with something other than  $\varepsilon$ . This technique also runs in linear time in the size of the NFA- $\varepsilon$  [13].  $\square$

**Almost closed languages** In analogy with Section 2.5, we can define a language  $L$  to be *almost prefix-closed* if there exists at least one, but only finitely many witnesses to the failure of the prefix-closed property. Analogously, we define *almost suffix-*, *almost factor-*, and *almost subword-closed*.

**Theorem 5.** *Let  $L$  be a regular language accepted by a DFA with  $n$  states. Then we can determine if  $L$  is almost prefix-closed (respectively, almost suffix-closed, almost factor-closed, almost subword-convex) in  $O(n^2)$  time.*

*Proof.* Just like the proof of Theorem 4.  $\square$

**Almost free languages** In a similar way, we can define a language  $L$  to be *almost prefix-free* if there exists at least one, but only finitely many witnesses to the failure of the prefix-free property. Analogously, we define *almost suffix-*, *almost factor-*, and *almost subword-free*.

**Theorem 6.** *Let  $L$  be a regular language accepted by a DFA with  $n$  states. Then we can determine if  $L$  is almost prefix-free (respectively, almost suffix-free, almost factor-free, almost subword-free) in  $O(n^2)$  time.*

*Proof.* Just like the proof of Theorem 4.  $\square$

### 3 Minimal witnesses

Recall that we let  $\trianglelefteq$  represent one of the four relations: factor, prefix, suffix, or subword. A necessary and sufficient condition that a language  $L$  be *not*  $\trianglelefteq$ -convex is the existence of a triple  $(u, v, w)$  of words, where  $u, w \in L$ ,  $v \notin L$ ,  $u \trianglelefteq v$ , and  $v \trianglelefteq w$ . We call such a triple a *witness* to the lack of  $\trianglelefteq$ -convexity. Let  $|x|$  denote the length of a string  $x$ . A witness  $(u, v, w)$  is *minimal* if every other witness  $(u', v', w')$  satisfies  $|w| < |w'|$ , or  $|w| = |w'|$  and  $|v| < |v'|$ , or  $|w| = |w'|$ ,  $|v| = |v'|$ , and  $|u| < |u'|$ . The *size* of a witness  $(u, v, w)$  is  $|w|$ .

Similarly, if  $L$  is not  $\trianglelefteq$ -closed, then  $(v, w)$  is a *witness* if  $w \in L$ ,  $v \notin L$ , and  $v \trianglelefteq w$ . A witness  $(v, w)$  is *minimal* if there exists no witness  $(v', w')$  such that  $|w'| < |w|$ , or  $|w'| = |w|$  and  $|v'| < |v|$ . The *size* is again  $|w|$ . For  $\trianglelefteq$ -freeness, *witness*, *minimal witness*, and *size* are defined as for  $\trianglelefteq$ -closure, except that both words are in  $L$ .

Suppose we are given a regular language  $L$  specified by an  $n$ -state DFA  $M$ , and we know that  $L$  is not  $\trianglelefteq$ -convex (respectively,  $\trianglelefteq$ -closed or  $\trianglelefteq$ -free). A natural question then is, what is a good upper bound on the size of the shortest witness that demonstrates the lack of this property?



### 3.1 Factor-convexity

From Theorem 1, we deduce the following corollary, which gives an  $O(n^3)$  upper bound for the length of a witness to the lack of factor-convexity.

**Corollary 3.** *Suppose  $L$  is accepted by a DFA with  $n$  states and  $L$  is not factor-convex. Then there exists a witness  $(u, v, w)$  such that  $|w| \leq 3n^3 + n^2 + n - 1$ .*

*Proof.* In our proof of Theorem 1, we constructed an NFA- $\varepsilon$   $M'$  with  $3n^3 + n^2 + n$  states accepting  $L(M')$ , the set of words  $w \in \Sigma^*$  such that there exist  $u, v \in \Sigma^*$  such that  $(u, v, w)$  is a witness. Thus, if  $M$  is not factor-convex,  $M'$  accepts such a word  $w$ , and the length of  $w$  is clearly bounded above by the number of states of  $M'$  minus 1.  $\square$

This bound is best possible up to a constant multiplicative factor, as stated in the following theorem.

**Theorem 7.** *There is a class of non-factor-convex regular languages  $L_n$ , accepted by DFA's with  $O(n)$  states, such the size of the minimal witness is  $\Omega(n^3)$ .*

The proof is postponed to Section 3.3 below.

Results analogous to Corollary 3 hold for prefix-, suffix-, and subword-convex languages. However, in some cases we can do better, as we show later.

**Factor-closure** Theorem 2 gives us a  $O(n^2)$  upper bound on the length of a witness to the failure of the factor-closed property:

**Corollary 4.** *If  $L$  is accepted by a DFA with  $n$  states and  $L$  is not factor-closed, then there exists a witness  $(v, w)$  such that  $|w| \leq 2n^2 + n - 1$ .*

This  $O(n^2)$  upper bound is best possible, up to a constant multiplicative factor. Let  $n \geq 3$  be an integer, and let  $M = (Q, \Sigma, \delta, q_0, F)$  be a DFA, where

$$Q = \{q_0, q_1, \dots, q_n, q_{n+1}, p_0, p_1, \dots, p_n, p_{n+1}\},$$

$\Sigma = \{0, 1\}$ , and  $F = Q \setminus \{q_{n+1}\}$ . The transition function is defined as follows:

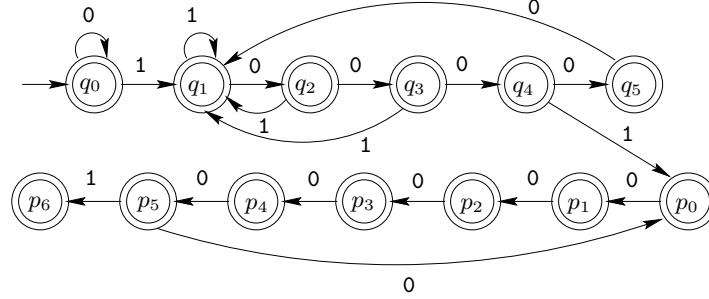
$$\delta(q_i, 0) = \begin{cases} q_0, & \text{if } i = 0; \\ q_{i+1}, & \text{if } 0 < i < n; \\ q_1, & \text{if } i = n; \\ q_{n+1}, & \text{if } i = n + 1; \end{cases}$$

$$\delta(q_i, 1) = \begin{cases} q_1, & \text{if } 0 \leq i < n - 1; \\ p_0, & \text{if } i = n - 1; \\ q_{n+1}, & \text{if } i \in \{n, n + 1\}; \end{cases}$$

$$\delta(p_j, 0) = \begin{cases} p_{j+1}, & \text{if } 0 \leq j < n; \\ p_0, & \text{if } j = n; \\ q_{n+1}, & \text{if } j = n + 1; \end{cases}$$

$$\delta(p_j, 1) = \begin{cases} q_{n+1}, & \text{if } 0 \leq j < n \text{ or } j = n + 1; \\ p_{n+1}, & \text{if } j = n. \end{cases}$$

The DFA  $M$  has  $2n + 4$  states. For  $n = 5$ ,  $M$  is illustrated in Figure 1.



**Fig. 1.** Example of the construction in Theorem 8 for  $n = 5$ . All unspecified transitions go to a rejecting “dead state”  $q_6$  (not shown) that cycles on all inputs.

Then we have the following theorem:

**Theorem 8.** *For the DFA  $M$  above, let  $L = L(M)$ . For any witness  $(u, v)$  to the lack of factor-closure we have  $|v| \geq (n+1)^2 - 1$ , and this bound is achievable.*

*Proof.* Let  $(u, v)$  be a minimal witness. Since the only rejecting state  $q_{n+1}$  in  $M$  leads only to itself, all the states along the accepting path of  $v$  are final. We claim that  $u$  is a suffix of  $v$ , that is,  $v = wu$  for some  $w$ . Otherwise, if the last letter of  $v$  is not the last letter of  $u$ , we can just omit it and get a shorter  $v$ , which

contradicts the minimality of  $v$ . Similarly, all the states along the rejecting path of  $u$  except the last one are final; otherwise, we get a shorter  $u$ .

First, we prove that the set of states along the accepting path of  $v$  includes both  $q$  states and  $p$  states. Let  $u = 0^i 1 u'$  for  $i \geq 0$ . Then  $\delta(q_1, u') = q_{n+1}$ . If  $\delta(q_0, w 0^i)$  is a  $p$  state, we are done. Otherwise, let  $\delta(q_0, w 0^i) = q_k$  for some  $0 \leq k \leq n$ . If  $k = n$ , then  $\delta(q_0, v) = \delta(q_0, w 0^i 1 u') = \delta(q_k, 1 u') = \delta(q_n, 1 u') = \delta(q_{n+1}, u') = q_{n+1}$ , a contradiction. If  $k = n - 1$ , then  $\delta(q_0, w 0^i 1) = \delta(q_k, 1) = p_0$ , which is a  $p$  state. Otherwise,  $\delta(q_0, v) = \delta(q_k, 1 u') = \delta(q_1, u') = q_{n+1}$ , a contradiction. Hence, the set of states along the accepting path of  $v$  includes both  $q$  states and  $p$  states.

We now prove that the set of states along the rejecting path of  $u$  includes only  $q$  states. Suppose it includes both  $q$  states and  $p$  states. Since there is only one transition from a  $q$  state to a  $p$  state and all transitions from a  $p$  state to a  $q$  state are to the rejecting state  $q_{n+1}$ , we have  $u = u_1 u_2$ , where  $\delta(q_0, u_1) = q_{n-1}$ , and

$$u_2 \in L_1 = 1(0^{n+1})^*(\varepsilon + 0 + 00 + \cdots + 0^{n-1})1.$$

Since  $u$  is a suffix of  $v$ , the last letter of  $v$  is also 1. So, by the construction of  $M$ , we have that  $v = v_1 v_2$ , where  $\delta(q_0, v_1) = q_{n-1}$ , and

$$v_2 \in L_2 = 1(0^{n+1})^* 0^n 1.$$

It is obvious that  $(\Sigma^* L_1) \cap (\Sigma^* L_2) = \emptyset$ , which contradicts the equality  $v_1 v_2 = v = w u = w u_1 u_2$ . Therefore, the set of states along the rejecting path of  $u$  includes only  $q$  states.

Consider the last block of 0's in the words  $u$  and  $v$ . By the structure of  $M$ , we have

$$u \in \Sigma^* 1(0^n)^* 0^{n-1} 1,$$

and

$$v \in \Sigma^* 1(0^{n+1})^* 0^n 1.$$

Therefore, the length of the last block of 0's is at least  $n(n+1) - 1$ . In other words,  $|u| \geq n(n+1) - 1 + 2 = n^2 + n + 1$ . Since the shortest word that leads to state  $q_{n-1}$  (which is the only state having a transition to a  $p$  state on input 1) is  $10^{n-2}$ , we also have  $|v| \geq 1 + n - 2 + n^2 + n + 1 = n^2 + 2n$ , and the first part of this theorem proved.

To see that equality is achieved, let  $u = 10^{n^2+n-1}1$  and  $v = 10^{n-2}u$ .  $\square$

**Factor-freeness** From Remark 1, we get

**Corollary 5.** *If  $L$  is accepted by a DFA with  $n$  states and  $L$  is not factor-free, then there exists a witness  $(v, w)$  such that  $|w| \leq 2n^2 + n - 1$ .*

Up to a constant multiplicative factor, Corollary 5 is best possible, as the following theorem shows.

**Theorem 9.** *There is a class of languages accepted by DFA's with  $O(n)$  states, such that the smallest witness to the lack of factor-freeness is of size  $\Omega(n^2)$ .*

*Proof.* Let

$$L = \mathbf{bb(a^n)^+b} \cup \mathbf{b(a^{n+1})^+b}.$$

This language can be accepted by a DFA with  $2n+6$  states. However, the shortest witness to the lack of factor-freeness is  $(\mathbf{ba}^{n(n+1)}\mathbf{b}, \mathbf{bba}^{n(n+1)}\mathbf{b})$ , which has size  $n^2 + n + 3$ .  $\square$

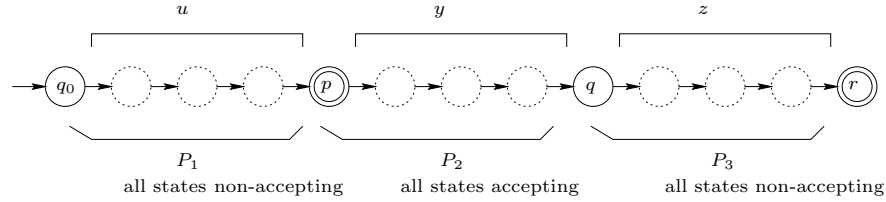
### 3.2 Prefix-convexity

For prefix-convexity, we have the following theorem.

**Theorem 10.** *Let  $M$  be a DFA with  $n$  states. If  $L(M)$  is not prefix-convex, there is a witness  $(u, v, w)$  with  $|w| \leq 2n - 1$ . Furthermore, this bound is best possible, as for all  $n \geq 2$ , there exists a unary DFA with  $n$  states that achieves this bound.*

*Proof.* If  $L(M)$  is not prefix-convex, then such a witness  $(u, v, w)$  exists. Without loss of generality, assume that  $(u, v, w)$  is minimal. Now write  $w = uyz$ , where  $v = uy$  and  $w = vz$ .

Let  $\delta(q_0, u) = p$ ,  $\delta(p, y) = q$ , and  $\delta(q, z) = r$ . Let  $P$  be the path from  $q_0$  to  $r$  traversed by  $uvw$ , and let  $P_1$  be the states from  $q_0$  to  $p$  (not including  $p$ ),  $P_2$  be the states from  $p$  to  $q$  (not including  $q$ ), and  $P_3$  be the states from  $q$  to  $r$  (not including  $r$ ); see Figure 2. Since  $(u, v, w)$  is minimal, we know that every state of  $P_3$  is rejecting, since we could have found a shorter  $w$  if there were an accepting state among them. Similarly, every state of  $P_2$  must be accepting, for, if there were a rejecting state among them, we could have found a shorter  $y$  and hence a shorter  $v$ . Finally, every state of  $P_1$  must be rejecting, since, if there were an accepting state, we could have found a shorter  $u$ .



**Fig. 2.** The acceptance path for  $w$

Let  $r_i = |P_i|$  for  $i = 1, 2, 3$ . There are no repeated states in  $P_3$ , for if there were, we could cut out the loop to get a shorter  $w$ ; the same holds for  $P_2$  and  $P_1$ . Thus  $r_i \leq n - 1$  for  $i = 1, 2, 3$ . Now  $P_1$  and  $P_2$  are disjoint, since all the states of  $P_1$  are rejecting, while all the states of  $P_2$  are accepting. Similarly, the states of  $P_3$  are disjoint from  $P_2$ . So  $r_1 + r_2 \leq n$  and  $r_2 + r_3 \leq n$ . It follows that  $r_1 + r_2 + r_3 \leq 2n - r_3$ . Since  $r_3 \geq 1$ , it follows that  $|w| \leq 2n - 1$ .

To see that  $2n - 1$  is optimal, consider the DFA of  $n$  states accepting the unary language

$$L = \mathbf{a}^{n-1}(\mathbf{a}^n)^*.$$

Then  $L$  is not prefix-convex, and the shortest witness is  $(\mathbf{a}^{n-1}, \mathbf{a}^n, \mathbf{a}^{2n-1})$ .  $\square$

**Prefix-closure** For prefix-closed languages we can get an even better bound.

**Theorem 11.** *Let  $M$  be an  $n$ -state DFA, and suppose  $L = L(M)$  is not prefix-closed. Then the minimal witness  $(v, w)$  showing  $L$  is not prefix-closed has  $|w| \leq n$ , and this is best possible.*

*Proof.* Assume that  $(v, w)$  is a minimal witness. Consider the path  $P$  from  $q_0$  to  $q = \delta(q_0, w)$ , passing through  $p = \delta(q_0, v)$ . Let  $P_1$  denote the part of the path  $P$  from  $q_0$  to  $p$  (not including  $p$ ) and  $P_2$ , the part of the path from  $p$  to  $q$  (not including  $q$ ). Then all the states traversed in  $P_2$  must be rejecting; otherwise, we would get a shorter  $w$ . Similarly, all the states traversed in  $P_1$  must be accepting, because otherwise we could get a shorter  $v$ . Neither  $P_1$  nor  $P_2$  contains a repeated state, because if they did, we could “cut out the loop” to get a shorter  $v$  or  $w$ . Furthermore, the states in  $P_1$  are disjoint from  $P_2$ . So the total number of states in the path to  $w$  (not counting  $q$ ) is at most  $n$ . Thus  $|w| \leq n$ .

The result is best possible, as the example of the unary language

$$L = (\mathbf{a}^n)^*$$

shows. This language is not prefix-closed, can be accepted by a DFA with  $n$  states, and the smallest witness is  $(\mathbf{a}, \mathbf{a}^n)$ .  $\square$

**Prefix-freeness** For the prefix-free property we have:

**Theorem 12.** *If  $L$  is accepted by a DFA with  $n$  states and is not prefix-free, then there exists a witness  $(v, w)$  with  $|w| \leq 2n - 1$ . The bound is best possible.*

*Proof.* The proof is similar to that of Theorem 10. The bound is achieved by a unary DFA accepting  $\mathbf{a}^{n-1}(\mathbf{a}^n)^*$ .  $\square$

### 3.3 Suffix-convexity

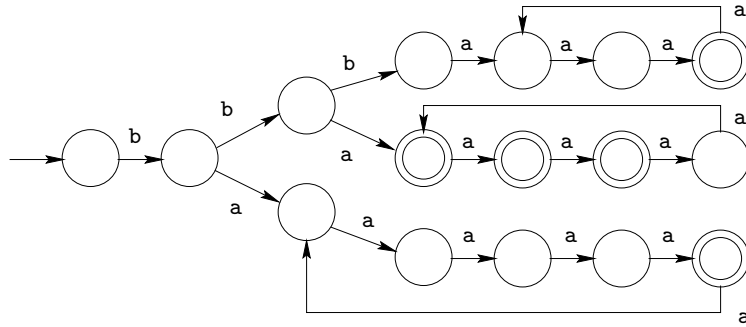
For the suffix-convex property, the cubic upper bound implied by Corollary 3 is best possible, up to a constant factor.

**Theorem 13.** *There is a class of non-suffix-convex regular languages  $L_n$ , accepted by DFA's with  $O(n)$  states, such the size of the minimal witness is  $\Omega(n^3)$ .*

*Proof.* Let

$$L = \text{bbb}(\text{a}^{n-1})^+ \cup \text{bb}(\text{a} + \text{aa} + \dots + \text{a}^{n-1})(\text{a}^n)^* \cup \text{b}(\text{a}^{n+1})^+.$$

Then  $L$  can be accepted by a DFA with  $3n + 5$  states, as illustrated in Figure 3.



**Fig. 3.** Example of the construction in Theorem 13 for  $n = 4$ . All unspecified transitions go to a rejecting “dead state”  $d$  that cycles on all inputs.

Suppose  $(u, v, w)$  is a witness; then  $w$  cannot be a word of the form  $\text{ba}^i$ , because no proper suffix of such a word is in  $L$ . Also,  $w$  cannot be a word of the form  $\text{bba}^i$ , because the only proper suffix in  $L$  is  $u = \text{ba}^i$ . But then there is no word  $v$  that lies strictly between  $u$  and  $w$  in the suffix order. So  $w$  must be of the form  $\text{bbba}^i$ . The only proper suffixes of  $w$  in  $L$  are of the form  $\text{bba}^i$  and  $\text{ba}^i$ . But we cannot have  $u = \text{bba}^i$  because, if we did, there would be no  $v$  strictly between  $u$  and  $w$  in the suffix order. So it must be that  $u = \text{ba}^i$ . Then the only word in  $\Sigma^*$  strictly between  $u$  and  $w$  in the suffix order is  $v = \text{bba}^i$ , and such a  $v$  is not in  $L$  if and only if  $i$  is a multiple of  $n$ . On the other hand, for  $u$  and  $w$  to be in  $L$ ,  $i$  must be a multiple of  $n + 1$  and  $n - 1$ , respectively.

It follows that  $L$  is not suffix-convex and the shortest witness is  $(\text{ba}^i, \text{bba}^i, \text{bbba}^i)$ , where  $i = \text{lcm}(n - 1, n, n + 1) \geq (n - 1)n(n + 1)/2$ .  $\square$

A similar technique can be used for non-factor-convex languages. This allows us to prove Theorem 7 in the same way Theorem 13, except we use the language  $Lb$  instead.

**Suffix-closure** Obviously, a witness to the failure of suffix-closure is also a witness to the failure of factor-closure. So the proof of Theorem 8 shows that the bound  $(n + 1)^2 - 1$  also holds for suffix-closed languages.

Ang and Brzozowski pointed out [2] that a language  $L$  is factor-closed if and only if  $L$  is both prefix-closed and suffix-closed. The next result shows that a long minimal witness for factor-closure must also be a witness for suffix-closure.

**Proposition 1.** *Let  $M$  be a DFA of  $n$  states, and  $L = L(M)$ . Let  $v$  be the shortest word such that there is  $u \notin L, v \in L, |v| > n$  and  $u$  is a factor of  $v$ . Then  $u$  is a suffix of  $v$ .*

*Proof.* Suppose  $u$  is not a suffix of  $v$ . Write  $v = v'a$  for  $a \in \Sigma$ . Then  $u$  is also a factor of  $v'$ . So,  $v' \notin L$ . Since  $|v'| \geq n$ , by the pumping lemma,  $v' = xyz$ , such that  $xz \notin L, |xz| < |xyz|$ . But  $xza \in L$  since  $xyzza = v'a = v \in L$ . This contradicts that  $v$  is the shortest.  $\square$

In other words, a long minimal witness for factor-closure must also be a witness for suffix-closure.

### Suffix-freeness

**Theorem 14.** *There exists a class of languages accepted by DFA's with  $O(n)$  states, such that the smallest witness to the lack of suffix-freeness is of size  $\Omega(n^2)$ .*

*Proof.* Let

$$L = \mathbf{bb}(\mathbf{a}^n)^+ \cup \mathbf{b}(\mathbf{a}^{n+1})^+.$$

This language is accepted by a DFA with  $2n + 5$  states. However, the shortest witness to the lack of suffix-freeness ( $\mathbf{ba}^{n(n+1)}, \mathbf{bba}^{n(n+1)}$ ) has size  $n^2 + n + 2$ .  $\square$

### 3.4 Subword-convexity

We now turn to subword properties. First, we recall some facts about the pumping lemma. If  $w = a_1 \cdots a_m$  with  $a_i \in \Sigma$  for  $1 \leq i \leq m$ , we write  $w[i, j]$  for the factor  $a_i \cdots a_j$ . Assume that  $M = (Q, \Sigma, \delta, q_0, F)$  is an  $n$ -state DFA,  $m \geq n$ , let  $q \in Q$ , and consider the state sequence

$$S(q, w) = (\delta(q, w[1, 0]), \dots, \delta(q, w[1, m])).$$

We know that some state in  $S(q, w)$  must appear more than once, because there are only  $n$  distinct states in  $M$ . Let  $\delta(q, w[1, i])$  be the first state that appears more than once in  $S$ , and let  $x = w[1, i]$ . Moreover, let  $\delta(q, w[1, j])$  be the first state in  $S(q, w)$  equal to  $\delta(q, w[1, i])$ , and let  $y = w[i + 1, j]$ . Finally, let  $z = w[j + 1, m]$ . Then  $w = xyz$ , where  $|xy| \leq n$ ,  $|y| > 0$ , and  $|z| \geq m - n$ , and  $\delta(q, x) = \delta(q, xy)$ . By the pumping lemma,  $xy^*z \subseteq L$ . By the definition of  $x$  and  $y$ , all the states in the sequence  $S(q, w[1, j - 1])$  are distinct. For a word  $w$  with  $|w| = m \geq n$ , we refer to the factorization  $w = xyz$  as the *canonical factorization of  $w$  with respect to  $q$* .

**Subword-closure** Here  $v \trianglelefteq w$  means  $v$  is a subword of  $w$ . If  $L = L(M)$  is not subword-closed, then  $(v, w)$  is a *witness* if  $w \in L, v \notin L$ , and  $v \trianglelefteq w$ .

**Lemma 1.** *Let  $M$  be a DFA with  $n \geq 2$  states such that  $L(M)$  is not subword-closed. For any witness  $(v, w)$ , there exists a witness  $(v', w')$  with  $|w'| \leq n$  and  $w' \trianglelefteq w$ .*

*Proof.* We will show that, for any witness  $(v, w)$  with  $|w| \geq n + 1$ , we can find a witness  $(v', w')$  with  $|w'| < |w|$  and  $w' \preceq w$ . The lemma then follows.

Suppose that  $(v, w)$  is a minimal witness, and  $|w| = m \geq n + 1$ . Then the canonical factorization of  $w$  is  $w = xyz$ , where  $|xy| \leq n$ ,  $|y| > 0$ , and  $|z| \geq m - n > 0$ .

If there is a  $z'$  such that  $z' \preceq z$  and  $xyz' \notin L$ , then  $xz' \notin L$ , since  $xyz'$  and  $xz'$  lead to the same state in  $M$ . Then  $(xz', xz)$  is a witness with  $|xz| < |w|$  and  $xz \preceq w$ . Thus we can assume that

$$z' \preceq z \text{ implies } xyz' \in L. \quad (1)$$

Since  $v \preceq w = xyz$ , we can write  $v = v_x v_y v_z$ , where  $v_x \preceq x$ ,  $v_y \preceq y$ , and  $v_z \preceq z$ . Clearly,  $v \preceq xyv_z$ . If  $v_z \neq z$ , then by (1), we have  $xyv_z \in L$ , and  $(v, xyv_z)$  is a witness with  $|xyv_z| < |w|$  and  $xyv_z \preceq w$ . Thus we may assume that our witness has the form  $(v_x v_y z, xyz)$ .

In the particular case that  $z' = \varepsilon$ , (1) implies that  $xy \in L$ . If  $y' \preceq y$  and  $xy' \notin L$ , then  $(xy', xy)$  is a witness with  $|xy'| < |w|$  and  $xy \preceq w$ . Thus

$$y' \preceq y \text{ implies } xy' \in L. \quad (2)$$

Finally, if  $x' \preceq x$  and  $x' \notin L$ , then  $(x', x)$  is a witness with  $|x| < |w|$  and  $x \preceq w$ . Thus

$$x' \preceq x \text{ implies } x' \in L. \quad (3)$$

Altogether, we may assume that all the states along the path spelling  $w$  in  $M$  are accepting. We know that the states in the sequence

$$S = (\delta(q_0, w[1, 0]), \dots, \delta(q_0, w[1, |xy| - 1]))$$

are all distinct. Also, the states in the sequence

$$S' = (\delta(q_0, v_x v_y z[1, 1]), \dots, \delta(q_0, v_x v_y z[1, |z| - 1]))$$

are all accepting and distinct; otherwise,  $v$  would not be shortest.

We now claim that no state can be in both  $S$  and  $S'$ . For suppose that  $\delta(q_0, w[1, i]) = \delta(q_0, v_x v_y z[1, k])$ , for some  $0 \leq i \leq |x|$ ,  $0 < k < |z|$ . Then  $(w[1, i]z[k + 1, |z|], xz)$  is a witness with  $|xz| < |w|$  and  $xz \preceq w$ , since  $w[1, i] = x[1, i]$ , and  $x[1, i]z[k + 1, |z|] \preceq xz$ . Next, if  $\delta(q_0, xy[1, j]) = \delta(q_0, v_x v_y z[1, k])$ , for some  $0 < j < |y|$ ,  $0 < k < |z|$ , then

$$(xy[1, j]z[k + 1, |z|], xyz[k + 1, |z|])$$

is a witness with  $|xyz[k + 1, |z|]| < |w|$  and  $xyz[k + 1, |z|] \preceq w$ , since  $xy[1, j]z[k + 1, |z|] \preceq xyz[k + 1, |z|]$ , and  $xyz[k + 1, |z|] \in L$  by (1).

Under these conditions  $M$  must have  $|xy| + (|z| - 1) = |xyz| - 1$  distinct accepting states, and at least one rejecting state. Hence  $|xyz| = |w| \leq n$  and we have found a witness with the required properties.  $\square$



**Corollary 6.** *Let  $M$  be a DFA with  $n \geq 2$  states. If  $L(M)$  is not subword-closed, there exists a witness  $(v, w)$  with  $|w| \leq n$ . Furthermore, this is the best possible bound, as there exists a unary DFA with  $n$  states that achieves this bound.*

*Proof.* If  $L$  is not subword-closed then it has a witness and, by Lemma 1, it has a witness  $(v, w)$  with  $|w| \leq n$ . This is the best possible bound for  $n \geq 2$ , since the language

$$(\mathbf{a}^n)^*(\varepsilon + \mathbf{a} + \cdots + \mathbf{a}^{n-2}),$$

accepted by a DFA with  $n$  states, has a minimal witness  $(\mathbf{a}^{n-1}, \mathbf{a}^n)$ .  $\square$

For  $n = 1$ ,  $L$  is either  $\emptyset$  or  $\Sigma^*$ , and these languages are subword-closed.

### Subword-freeness

**Lemma 2.** *Let  $M$  be a DFA with  $n \geq 2$  states such that  $L(M)$  is not subword-free. For any witness  $(u, w)$ , there exists a witness  $(u', w')$  with  $|w'| \leq 2n - 1$ , and  $w' \preceq w$ .*

*Proof.* We will show that, for any witness  $(u, w)$  with  $|w| \geq 2n$ , we can find a witness  $(u', w')$  with  $|w'| < |w|$  and  $w' \preceq w$ . The lemma then follows.

Let the canonical factorization of  $w$  with respect to  $q_0$  be  $w = xyz$ , where  $|xy| \leq n$ ,  $|y| > 0$ , and  $|z| \geq n > 0$ . Then we also have a canonical factorization of  $z = x'y'z'$  with respect to state  $q = \delta(q_0, xy)$ , where  $|x'y'| \leq n$ ,  $|y'| > 0$ , and  $|z'| \geq 0$ . Now we have a witness  $(xx'z', xx'y'z') = (xx'z', xz)$  with  $|xz| < |w|$  and  $xz \preceq w$ .  $\square$

**Corollary 7.** *Let  $M$  be a DFA with  $n \geq 2$  states. If  $L(M)$  is not subword-free, there exists a witness  $(u, w)$  with  $|w| \leq 2n - 1$ . This is the best possible bound, as there exists a unary DFA with  $2n - 1$  states that achieves this bound.*

*Proof.* If  $L$  is not subword-free then it has a witness and, by Lemma 2, it has a witness  $(v, w)$  with  $|w| \leq 2n - 1$ . This is the best possible bound for  $n \geq 2$ , since the language

$$\mathbf{a}^{n-1}(\mathbf{a}^n)^*,$$

accepted by a DFA with  $n$  states, has a minimal witness  $(\mathbf{a}^{n-1}, \mathbf{a}^{2n-1})$ .  $\square$

For  $n = 1$ ,  $L$  is either  $\emptyset$  or  $\Sigma^*$ . Only  $\Sigma^*$  is not subword-free, and has a minimal witness  $(\varepsilon, \mathbf{a})$  for any  $a \in \Sigma$ .

### Subword-convexity

**Lemma 3.** *Let  $M$  be a DFA with  $n \geq 2$  states such that  $L(M)$  is not subword-convex. For any witness  $(u, v, w)$ , there exists a witness  $(u', v', w')$  with  $w' \preceq w$ , and  $|w'| \leq 3n - 2$ .*

*Proof.* We will show that, for any witness  $(u, v, w)$  with  $|w| \geq 3n - 1$ , we can find a witness  $(u', v', w')$  with  $|w'| < |w|$  and  $w' \sqsubseteq w$ . The lemma then follows.

We assume without loss of generality that  $v$  is a shortest possible word corresponding to the given  $w$ , and  $u$  is a shortest word corresponding to  $v$  and  $w$ .

First, consider the witness  $(u, v)$  to the lack of subword-closure of the language  $\bar{L}$ . By Lemma 1, there exists a witness  $(u', v')$  to the failure of subword-closure of  $\bar{L}$  such that  $v' \sqsubseteq v$  and  $|v'| \leq n$ . Therefore we can assume that we have a witness  $(u, v, w)$  to the failure of subword-convexity such that  $|v| \leq n$ .

Suppose that  $(u, v, w)$  is a minimal witness, and  $|w| \geq 3n - 1$ . Then the canonical factorization of  $w$  is  $w = x_1 y_1 z_1$ , where  $|x_1 y_1| \leq n$ ,  $|y_1| > 0$ , and  $|z_1| \geq 2n - 1 \geq n > 0$ . Consider the states

$$p_0 = \delta(q_0, x_1 y_1), p_1 = \delta(q_0, x_1 y_1 z_1[1, 1]), \dots, p_{|z_1|} = \delta(q_0, x_1 y_1 z_1).$$

Since  $|z_1| \geq n$ , there must be at least one pair  $(p_i, p_j)$  of states such that  $p_i = p_j$ . If  $p_0$  is the state that is repeated, let  $i$  be the greatest index such that  $p_0 = p_i$ , and let  $x_2 = \varepsilon$ ,  $y_2 = z_1[1, i]$ , and  $z_2 = z_1[i + 1, |z_1|]$ . If  $p_i$  is the first state that is repeated, let  $j$  be the greatest index such that  $p_i = p_j$ , and let  $x_2 = z_1[1, i]$ ,  $y_2 = z_1[i + 1, j]$ , and  $z_2 = z_1[j + 1, |z_1|]$ . If  $\delta(q_0, x_1 y_1 x_2 y_2), \delta(q_0, x_1 y_1 x_2 y_2 z_2[1, 1]), \dots, \delta(q_0, x_1 y_1 x_2 y_2 z_2)$  has no repeated states, we stop. Otherwise, we apply the same procedure to  $z_2$ , and so on. In any case, eventually we reach a  $z_k$  for which no repeated states exist. Then we have the factorization

$$w = x_1 y_1 x_2 y_2 \cdots x_k y_k z_k,$$

where  $x_1 y_1^* x_2 y_2^* \cdots x_k y_k^* z_k \subseteq L$ ,  $|x_2 \cdots x_k z_k| < n$  (otherwise, there would be repeated states),  $|y_i| > 0$ , for  $i = 1, \dots, k$ , and  $k \geq 2$ .

For any  $y'_2 \sqsubseteq y_2, \dots, y'_k \sqsubseteq y_k$ , we have  $x_1 y_1 x_2 y'_2 \cdots x_k y'_k z_k \in L$ . Otherwise, the triple

$$(x_1 x_2 \cdots x_k z_k, x_1 x_2 y'_2 \cdots x_k y'_k z_k, x_1 x_2 y_2 \cdots x_k y_k z_k)$$

is a witness with  $|x_1 x_2 y_2 \cdots x_k y_k z_k| < |w|$ , and  $x_1 x_2 y_2 \cdots x_k y_k z_k \sqsubseteq w$ .

Since  $v \sqsubseteq w$ , we can now write  $v = v_{x_1} v_{y_1} v_{x_2} v_{y_2} \cdots v_{x_k} v_{y_k} v_{z_k}$ , where  $v_{x_1} \sqsubseteq x_1$ , etc. If there is a  $y_i$  with  $i \geq 2$ , such that  $v_{y_i} = \varepsilon$ , then we can replace that  $y_i$  by  $\varepsilon$  in  $w$  and obtain a smaller witness. Hence each  $v_{y_i}$  must be nonempty. By the same argument, if there is a letter in  $y_i$ , for  $i \geq 2$ , that is not used in  $v_{y_i}$ , then that letter can be removed, yielding a smaller witness. Therefore  $y_i = v_{y_i}$  for  $i = 2, \dots, k$ . We claim that  $|y_2 \cdots y_k| < |v|$ ; otherwise  $v = v_{y_2} \cdots v_{y_k} = y_2 \cdots y_k$  and  $(u, v, x_1 x_2 y_2 \cdots x_k y_k z_k)$  is a witness with  $|x_1 x_2 y_2 \cdots x_k y_k z_k| < |w|$ . Thus  $|y_2 \cdots y_k| < |v| \leq n$ , and

$$|w| = |x_1 y_1| + |x_2 \cdots x_k z_k| + |y_2 \cdots y_k| \leq n + (n - 1) + (n - 1) = 3n - 2. \quad \square$$

**Corollary 8.** *Let  $M$  be a DFA with  $n \geq 2$  states. If  $L(M)$  is not subword-convex, there exists a witness  $(u, v, w)$  with  $|w| \leq 3n - 2$ .*

We do not know whether  $3n - 2$  is the best bound. The unary language

$$\mathbf{a}^{n-1}(\mathbf{a}^n)^*$$

is accepted by a DFA with  $n$  states and has a minimal witness  $(\mathbf{a}^{n-1}, \mathbf{a}^n, \mathbf{a}^{2n-1})$ , showing that  $2n - 1$  is achievable.

## 4 NL-completeness for DFA's

Let NL denote the class of problems solvable in nondeterministic logarithmic space. As two referees of an earlier version of this paper remarked, we can also prove

**Theorem 15.** *Testing each of the prefix-free, prefix-closed, and prefix-convex properties is NL-complete, and the same holds for suffixes, factors, and subwords.*

*Proof.* We illustrate the idea for the prefix-free property, using a reduction from a variant of GAP (the graph accessibility problem) [14]: given a directed acyclic graph  $G = (V, E)$  on the vertices  $\{q_1, q_2, \dots, q_n\}$ , is there a path from  $q_1$  to  $q_n$ ?

Given such a directed graph  $G$ , we discard all edges into  $q_1$  and out of  $q_n$  to obtain a new graph  $G' = (V', E')$ . Then we create a DFA  $M = (Q, \Sigma, \delta, q_0, F)$  as follows:

- $Q = V' \cup \{q_0, q_{n+1}, d\}$ ;
- $\Sigma = E' \cup \{\#, \$\}$ ;
- $F = \{q_1, q_{n+1}\}$ ;
- $\delta(q_0, \#) = q_1$ ;
- $\delta(q_n, \$) = q_{n+1}$ ;
- $\delta(q_i, e) = q_j$  if there is an edge  $e$  from  $q_i$  to  $q_j$  in  $G'$ ;
- $\delta(q_i, c) = d$  for all previously undefined transitions on symbols  $c \in \Sigma$ ;
- $\delta(d, c) = d$  for all  $c \in \Sigma$ .

It is clear that this transformation can be done in logarithmic space.

We claim that  $L(M)$  fails to be prefix-free if and only if there is a directed path from  $q_1$  to  $q_n$  in  $G$ . Suppose there is such a path, with edges  $e_1, e_2, \dots, e_m$ ; then there is a path that uses no edges into  $q_1$  or out of  $q_n$ . In  $M$  this corresponds to the word  $\#e_1e_2 \dots e_m\$$ ; since the word  $\#$  is also accepted, this means that  $L(M)$  is not prefix-free.

On the other hand, by the construction,  $M$  accepts  $\#$  and possibly words of the form  $\#e_1e_2 \dots e_m\$$ . Since  $M$  accepts  $\#$  by construction,  $L(M)$  fails to be prefix-free if and only if it accepts a word of the form  $\#e_1e_2 \dots e_m\$$ . But then  $e_1, e_2, \dots, e_m$  is a path connecting  $q_1$  to  $q_n$ .

Finally, testing prefix-freeness is in NL because we can (implicitly) construct the NFA as in Section 2.1 and nondeterministically test connectivity between  $q_0$  and a final state.

The same idea can be used to prove the results for the other 11 possibilities.

## 5 Languages specified by other means

In this section we reconsider some of the same problems, this time assuming that the given language is specified by an NFA or a context-free grammar, instead of a DFA.

## 5.1 Languages specified by NFA's

Some of our decision problems become PSPACE-complete if  $M$  is represented by an NFA. Our fundamental tool is the following classical lemma [15]:

**Lemma 4.** *Let  $T$  be a one-tape deterministic Turing machine and  $p(n)$  a polynomial such that  $T$  never uses more than  $p(|x|)$  space on input  $x$ . Then there is a finite alphabet  $\Delta$  and a polynomial  $q(n)$  such that we can construct a regular expression  $r_x$  in  $q(|x|)$  steps, such that  $L(r_x) = \Delta^*$  if  $T$  doesn't accept  $x$ , and  $L(r_x) = \Delta^* - \{w\}$  for some nonempty  $w$  (depending on  $x$ ) otherwise. Similarly, we can construct an NFA  $M_x$  in  $q(|x|)$  steps, such that  $L(M_x) = \Delta^*$  if  $T$  doesn't accept  $x$ , and  $L(M_x) = \Delta^* - \{w\}$  for some nonempty  $w$  (depending on  $x$ ) otherwise.*

**Theorem 16.** *The problem of deciding whether a given regular language  $L$ , represented by an NFA or regular expression, is prefix-convex (respectively, suffix-, factor-, subword-convex), or prefix-closed (respectively, suffix-, factor-, subword-closed) is PSPACE-complete.*

*Proof.* We prove the result for factor-convexity, the other results being proved in the same way.

First, we show that the problem of deciding factor-convexity is in PSPACE. We actually show that we can solve it in NSPACE, and then use Savitch's theorem that PSPACE = NSPACE.

Suppose  $L$  is accepted by an NFA  $M$  with  $n$  states. Then, by the subset construction,  $L$  is accepted by a DFA with  $\leq 2^n$  states. From Theorem 1 above, we see that if  $L$  is not factor-convex, we can demonstrate this by exhibiting  $u, v, w$  with  $u$  a prefix of  $v$  and  $v$  is a prefix of  $w$ , and  $u, w \in L$ ,  $v \notin L$  and then checking that these conditions are fulfilled. Furthermore, from Corollary 3, if such  $u, v, w$  exist, then  $|u|, |v|, |w| = O((2^n)^3)$ . In polynomial space, we can count up to  $2^{3n}$ . Write  $w = x_1x_2x_3x_4x_5$ , and let  $v = x_2x_3x_4$  and  $u = x_3$ . We use boolean matrices to keep track of, for each state of  $M$ , what state we would be in after reading prefixes of  $w$ . We guess the appropriate words  $x_1, x_2, x_3, x_4, x_5$  symbol-by-symbol, using a counter to ensure these words are shorter than  $2^{3n}$ . We then verify that  $x_3$  and  $x_1x_2x_3x_4x_5$  are in  $L$  and  $x_2x_3x_4$  is not.

Now we show that the problem is PSPACE-hard. Since  $\Delta^*$  is factor-convex and  $\Delta^* - \{w\}$  is not if  $w \neq \varepsilon$ , we could use an algorithm solving the factor-convex problem to solve decidability for polynomial-space bounded Turing machines.  $\square$

For the prefix-, suffix-, and factor-closed properties, PSPACE-hardness was already proved by Hunt and Rosenkrantz [16, Thm. 3.4]; also see [12].

The situation is different for deciding the property of prefix-freeness, suffix-freeness, etc., for languages represented by NFA's, as the following theorem shows. This was proved by Han et al. [10] through a different approach.

**Theorem 17.** *Let  $M$  be an NFA with  $n$  states and  $t$  transitions. Then we can decide in  $O(n^2 + t^2)$  time whether  $L(M)$  is prefix-free (respectively, suffix-free, factor-free, subword-free).*

*Proof.* We give the full details for prefix-freeness, and sketch the result for the other three cases.

Given  $M = (Q, \Sigma, \delta, q_0, F)$ , create an NFA  $M'$  accepting  $L(M)\Sigma^+$ . This can be done, for example, by adding a transition on each  $a \in \Sigma$  from each old final state of  $M$  to a new state  $q_f$ , and having a loop on  $q_f$  to itself on each  $a \in \Sigma$ . Finally, let the new set of final states for  $M'$  be  $\{q_f\}$ . Clearly,  $L(M)$  is prefix-free if and only if  $L(M) \cap L(M') = \emptyset$ . We can construct an NFA  $M''$  accepting  $L(M) \cap L(M')$  using the usual “direct product” construction. If the original  $M$  had  $n$  states and  $t$  transitions, the new  $M'$  has  $n + 1$  states and at most  $t + 2n|\Sigma|$  transitions. So  $M''$  has  $n(n + 1)$  states and at most  $t(t + 2n|\Sigma|)$  transitions. Without loss of generality we can assume that  $t \geq n - 1$  (otherwise  $M$  is not connected), and hence it costs  $O(n^2 + t^2)$  to check whether  $L(M'') = \emptyset$  using depth-first search.

For suffix-freeness, we carry out a similar construction for  $L(M) \cap \Sigma^+L(M)$ . For factor-freeness, we carry out a similar construction for  $L(M) \cap (\Sigma^+L(M)\Sigma^* \cup \Sigma^*L(M)\Sigma^+)$ .

For subword-freeness, we carry out a similar but slightly more involved construction, which is as follows: create  $M'$  by making two copies of  $M$ . Add a transition from each state  $q$  to its copy  $q'$  on each letter of  $\Sigma$ , and add transitions from each copy  $q'$  to itself on each letter of  $\Sigma$ . The final states of  $M'$  are the final states in the part corresponding to the copied states. Formally,  $M' = (Q \cup Q', \Sigma, \delta, q, F')$  where  $Q' = \{q' : q \in Q\}$ ,  $F' = \{q' : q \in F\}$ , and  $\delta'(q, a) = \delta(q, a) \cup \{q'\}$  for all  $q \in Q$ ,  $a \in \Sigma$ , and  $\delta'(q', a) = \delta(q, a)' \cup \{q'\}$  for all  $q \in Q$ ,  $a \in \Sigma$ . Then  $M'$  accepts the language of all words that are strict superwords of words accepted by  $M$ . We now create the NFA for  $L(M) \cap L(M')$  as before.  $\square$

**Minimal witnesses for NFA’s** We have already seen that the length of the minimal witness for the lack of convexity or closure is polynomial in the size of the DFA. For the case of NFA’s, however, this bound no longer holds.

**Theorem 18.** *There is a class of NFA’s with  $O(n)$  states such that the shortest witness to the lack prefix-convexity (respectively, suffix-, factor-, subword-convexity) or prefix-closure (respectively, suffix-, factor-, subword-closure) is of length  $2^{\Omega(n)}$ .*

*Proof.* In Ellul et al. [17, §5, p. 433] the authors show how to construct a regular expression  $E$  of length  $O(n)$  that accepts all words up to some length  $2^{\Omega(n)}$ , at which point a word is omitted. From  $E$  one can construct an NFA with  $O(n)$  states accepting an  $L$  with the desired property.  $\square$

For the prefix-free, etc., properties, we have

**Theorem 19.** *There exists a class of languages, accepted by NFA’s with  $O(n)$  states and  $O(n)$  transitions, such that the minimal witness for the lack of prefix-freeness is of length  $\Omega(n^2)$ .*

*Proof.* For non-prefix-freeness, we can use the reverse of the language defined in the proof of Theorem 14.  $\square$

For the lack of subword-freeness, we cannot improve the bound we obtained for DFA's in Corollary 7, as the proof we presented there also works for NFA's.

## 5.2 Languages specified by context-free grammars

If  $L$  is represented by a context-free grammar, then the decision problems corresponding to convex and closed languages become undecidable. This follows easily from a well-known result that the set of invalid computations of a Turing machine is a CFL [18, Lemma 8.7, p. 203].

Similarly, the decision problems corresponding to the properties of prefix-free, suffix-free, and factor-free become undecidable for CFL's, as shown by Jürgensen and Konstantinidis [5, Thm. 9.5, p. 581].

However, testing subword-freeness is still decidable for CFL's:

**Theorem 20.** *There is an algorithm that, given a context-free grammar  $G$ , will decide if  $L(G)$  is subword-free.*

*Proof.* If  $L = L(G)$  is infinite, then  $L$  is not subword-free by the pumping lemma. For if  $|w|$  is sufficiently large, then we can factor  $w$  as  $uvxyz$ , where  $|vy| \geq 1$ , such that  $uxz \in L$ . But  $uxz$  is a subword of  $w$ . We can test if  $L(G)$  is infinite by a well-known result [18, Thm. 6.6, p. 137]. Otherwise, if  $L(G)$  is finite, we can enumerate all its words and test each for the subword-free property.  $\square$

We can also consider the shortest witness to the failure of the subword-free property.

**Theorem 21.** *Let  $G$  be a context-free grammar in Chomsky normal form with  $n$  variables. If  $L(G)$  fails to be subword-free, then there is a witness  $(u, v)$  with  $u, v \in L(G)$  and  $u$  a subword of  $v$  such that  $|v| \leq 2^{n-1}$ . Furthermore, for each  $n$ , there exists a context-free grammar in Chomsky normal form with  $n$  variables, and of size  $O(n)$ , such that the shortest witness to the failure of the subword-free property is of length  $2^{n-1}$ .*

*Proof.* An easy induction shows that if  $G$  has  $n$  variables and  $L(G)$  is finite, then the longest word in  $L(G)$  is of length  $\leq 2^{n-1}$ .

It is easy to construct a Chomsky normal form grammar for the language  $\{a, a^{2^{n-1}}\}$  using  $n$  variables and  $n + 1$  productions.

## 6 Conclusions

We have shown that we can decide in  $O(n^3)$  time whether a language specified by a DFA is prefix-, suffix-, factor-, or subword-convex, and that the corresponding closure and freeness properties can be tested in  $O(n^2)$  time. If  $L$  is specified by an NFA or a regular expression, these problems are PSPACE-complete.

Our results about the sizes of minimal witnesses for the various classes are summarized in Table 1. All results are known to be best possible, except the  $3n - 2$  upper bound for subword-convexity; in this case, we do not know whether the bound is achievable.

**Table 1.** Sizes of witnesses

property relation	convexity	closure	freeness
factor	$\Theta(n^3)$	$\Theta(n^2)$	$\Theta(n^2)$
prefix	$2n - 1$	$n$	$2n - 1$
suffix	$\Theta(n^3)$	$\Theta(n^2)$	$\Theta(n^2)$
subword	$3n - 2$	$n$	$2n - 1$

## 7 Acknowledgments

This research was supported by the Natural Sciences and Engineering Research Council of Canada. We thank the referees for their helpful comments, especially about NL-completeness.

## References

1. Thierrin, G.: Convex languages. In Nivat, M., ed.: Automata, Languages, and Programming. North-Holland (1973) 481–492
2. Ang, T., Brzozowski, J.: Continuous languages. In Csuhaaj-Varjú, E., Ésik, Z., eds.: Proc. 12th International Conference on Automata and Formal Languages. Computer and Automation Research Institute, Hungarian Academy of Sciences (2008) 74–85
3. Ang, T., Brzozowski, J.: Languages convex with respect to binary relations, and their closure properties. *Acta Cybernetica*, to appear (2009)
4. Berstel, J., Perrin, D.: Theory of Codes. Academic Press, New York (1985)
5. Jürgensen, H., Konstantinidis, S.: Codes. In Rozenberg, G., Salomaa, A., eds.: Handbook of Formal Languages, Vol. 1. Springer-Verlag (1997) 511–607
6. de Luca, A., Varricchio, S.: Some combinatorial properties of factorial languages. In Capocelli, R., ed.: Sequences. Springer (1990) 258–266
7. Brzozowski, J., Shallit, J., Xu, Z.: Decision problems for convex languages. In Dediu, A.H., Ionescu, A.M., Martín-Vide, C., eds.: LATA 2009. Volume 5457 of Lect. Notes Comp. Sci. Springer (2009) 247–258
8. Brzozowski, J.A., Shallit, J., Xu, Z.: Decision problems for convex languages. Preprint, available at <http://arxiv.org/abs/0808.1928> (2008)
9. Béal, M.P., Crochemore, M., Mignosi, F., Restivo, A., Sciortino, M.: Computing forbidden words of regular languages. *Fund. Inform.* **56** (2003) 121–135
10. Han, Y.S., Wang, Y., Wood, D.: Infix-free regular expressions and languages. *Internat. J. Found. Comp. Sci.* **17** (2006) 379–393

11. Gill, A., Kou, L.: Multiple-entry finite automata. *J. Comput. System. Sci.* **9** (1974) 1–19
12. Kao, J.Y., Rampersad, N., Shallit, J.: On NFAs where all states are final, initial, or both. Preprint, available at <http://arxiv.org/abs/0808.2417> (2009)
13. Tarjan, R.: Depth-first search and linear graph algorithms. *SIAM J. Comput.* **1** (1972) 146–160
14. Jones, N.D.: Space-bounded reducibility among combinatorial problems. *J. Comput. System Sci.* **11** (1975) 68–85
15. Aho, A., Hopcroft, J., Ullman, J.: *The Design and Analysis of Computer Algorithms*. Addison-Wesley (1974)
16. Hunt III, H.B., Rosenkrantz, D.J.: Computational parallels between the regular and context-free languages. *SIAM J. Comput.* **7** (1978) 99–114
17. Ellul, K., Krawetz, B., Shallit, J., Wang, M.-w.: Regular expressions: new results and open problems. *J. Automata, Languages, and Combinatorics* **10** (2005) 407–437
18. Hopcroft, J.E., Ullman, J.D.: *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley (1979)