Information and Computation 209 (2011) 353–367

Contents lists available at ScienceDirect

Information and Computation

journal homepage: www.elsevier.com/locate/ic

Decision problems for convex languages

Janusz Brzozowski *, Jeffrey Shallit, Zhi Xu [1]

David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, ON, Canada N2L 3G1

ARTICLE INFO

Article history:
Available online 18 November 2010

Keywords:
Finite automaton
Complexity
Convex language
Regular language
Prefix-free
Suffix-free
Ideal

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 $w_0, w_1, \ldots, w_n, x_1, x_2, \ldots, x_n$ 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. 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. 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.

A language $L \subseteq \Sigma^*$ is a *right ideal* (respectively, *left ideal*, *two-sided ideal*, *all-sided ideal*) if it is non-empty and satisfies $L = L\Sigma^*$ (respectively, $L = \Sigma^*L$, $L = \Sigma^*L\Sigma^*$, $L = \Sigma^* \sqcup L$, where $\sqcup$ is the shuffle operator). Ideals and closed languages are related as follows [1,2]: a non-empty language is a right ideal (respectively, left, two-sided, or all-sided ideal) if and only if its complement is prefix-closed (respectively, suffix-, factor-, or subword-closed). Furthermore, ideals and converse-closed languages coincide in the following sense: A non-empty language is a right (respectively, left, two-sided, or all-sided) ideal if and only if it is converse-prefix-closed (respectively, converse-suffix-closed, converse-factor-closed, or converse-subword-closed).

* Corresponding author.
E-mail addresses: brzozo@uwaterloo.ca (J. Brzozowski), shallit@cs.uwaterloo.ca (J. Shallit), zhi_xu@uwo.ca (Z. Xu).
[1] Present address: University of Western Ontario, Department of Computer Science, London, ON, Canada N6A 5B7.

0890-5401/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ic.2010.11.009

The convex, free, closed and converse-closed classes contain many interesting languages that have received considerable attention. We now give some examples of papers on this subject; the list is by no means exhaustive.

The concept of a language convex with respect to the subword relation was introduced by Thierrin [3] in 1973. Recently Ang and Brzozowski [1,2] generalized the concept of convex languages to arbitrary relations, and in particular to the prefix, suffix, and factor relations.

Subword-free languages were considered by Haines [4] in 1969. In 1991, Jürgensen and Yu [5], studied languages that are independent with respect to binary relations. These languages include hypercodes [6,7] which are subword-free, infix codes [7] which are factor-free, and prefix and suffix codes [8] which are prefix- and suffix-free, respectively. For more information about codes, see [8,9]. In 2001 Jürgensen et al. [10] examined languages defined by arbitrary relations on an arbitrary set, and then specialized their results to the free monoid Σ^* generated by an alphabet Σ. They continued the study of languages independent with respect to binary relations — that is, free languages in our terminology.

Subword-closed languages were studied in 1969 by Haines [4], and also in 1973 by Thierrin [3]. Suffix-closed languages were considered by Gill and Kou [11] in 1974, later also in [12,13], and more recently in [14,15]. Factor-closed languages, often called *factorial*, were studied by de Luca and Varricchio [16].

Left and right ideals were studied by Paz and Peleg [17] in 1965 under the names "ultimate definite" and "reverse ultimate definite events". All-sided ideals were used by Haines [4] (not under that name) in 1969 in connection with subword-free and subword-closed languages, and by Thierrin [3] in 1973 in connection with subword-convex languages. De Luca and Varricchio [16] showed in 1990 that a language is factor-closed (also called "factorial") if and only if it is the complement of a two-sided ideal. Converse-subword-closed languages also appear under the names ϱ_h-ideal languages in a 2001 paper of Jürgensen et al. [10]. In 2001 Shyr [7] studied right, left, and two-sided ideals and their generators in connection with codes. It was noted in [1,2] 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 Σ^*.

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., right ideal, left ideal, etc., for a total of 16 different problems. However, the problem of testing whether a language is a right ideal, left ideal, etc., is equivalent to testing whether its complement is prefix-closed, suffix-closed, etc. Hence the number of problems reduces to 12. 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 (which we call *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.

Earlier, shorter versions of this paper appeared as a preprint [18] and in the LATA 2009 conference [19].

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

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

Then $L(M') = \emptyset$ if and only if $L(M)$ is $\trianglelefteq$-convex. We can test the emptiness of $L(M')$ using depth-first search in time linear in the size of M' (see [20], for example). This gives an $O(n^3)$ algorithm for testing $\trianglelefteq$-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-ε 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 $Q' = Q \times Q \times Q \times \{1, 2, 3, 4, 5\}$, $q_0' = [q_0, q_0, q_0, 1]$, $F' = F \times (Q \setminus F) \times F \times \{5\}$, 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-ε M' has $3n^3 + n^2 + n$ reachable states and $(3|\Sigma| + 2)n^3 + (|\Sigma| + 1)(n^2 + n)$ transitions, where $|\Sigma|$ is the cardinality of Σ.

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 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_α 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 ε-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 ε-move. Next, we process the symbols of u using Rule 5, ending in the state $[q_3, q_b, q_\alpha, 3]$. Then we use Rule 6 to move to $[q_3, q_b, q_\alpha, 4]$ by an ε-move. Next, we process the symbols of u'' using Rule 7, ending in the state $[q_4, q_c, q_\alpha, 4]$. Then we use Rule 8 to move to $[q_4, q_c, q_\alpha, 5]$ by an ε-move. Finally, we process the symbols of v'' using Rule 9, ending in the state $[q_5, q_c, q_\alpha, 5]$, 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. Thus, we have proved the following theorem:

Theorem 1. *If M is a DFA with n states, there exists an NFA-ε M' with $3n^3 + n^2 + n$ states and $(3|\Sigma| + 2)n^3 + (|\Sigma| + 1)(n^2 + n)$ transitions that accepts the language $L(M') = \{w \in \Sigma^* : w \in L$ and there exist $u, v \in \Sigma^*$ such that $u \in L$, $v \notin L$, u is a factor of v, and v is a factor of $w\}$.*

Corollary 2. *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'. □

2.1.1. 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-ε M' that accepts the language

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

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 $Q' = Q \times Q \times \{1, 2, 3\}$; $q_0' = [q_0, q_0, 1]$; $F' = F \times (Q \setminus F) \times \{3\}$; and δ' is defined as follows:

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 $p, q \in Q, a \in \Sigma$.

The NFA M' has $2n^2 + n$ reachable states and $(2|\Sigma| + 1)n^2 + (|\Sigma| + 1)n$ transitions. Thus we have the following result, which was previously obtained by Béal et al. [21, Proposition 5.1, p. 13] through a slightly different approach:

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

2.1.2. Converse-factor-closed languages

The converse of the relation "u is a factor of v" is "v contains u as a factor". 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 $\Sigma^* \setminus L$ is factor-closed. Then the following is an immediate consequence of Theorem 3:

Corollary 4. *We can decide if a given regular language L accepted by a DFA with n states is a two-sided ideal in $O(n^2)$ time.*

The results above apply to other converse-closed languages.

2.1.3. Factor-free languages

Factor-free (also known as infix-free) languages have been studied recently by Han et al. [22], who gave efficient algorithms for determining if the language accepted by an NFA is prefix-, suffix-, or factor-free.

Remark 5. 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 $F' = F \times F \times \{3\}$. As in the factor-closed case, the DFA has $2n^2 + n$ states.

Remark 6. Similar results hold for prefix-, suffix-, 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 $Q' = Q \times Q \times Q \times \{1, 2, 3\}$, $q'_0 = [q_0, q_0, q_0, 1]$, $F' = F \times (Q \setminus F) \times F \times \{3\}$, 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 $3n^3$ reachable states and $(3|\Sigma| + 2)n^3$ transitions.

2.2.1. Prefix-closed languages

By varying the construction in Section 2.1, we have

Theorem 7. *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 [11,14,15].

2.2.2. Prefix-free languages

Already addressed in Remark 6.

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 $Q' = Q \times \{q_0\} \times \{q_0\} \times \{1\} \cup Q \times Q \times \{q_0\} \times \{2\} \cup Q \times Q \times Q \times \{3\}$, $q'_0 = [q_0, q_0, q_0, 1]\}$, $F' = F \times (Q \setminus F) \times F \times \{3\}$, and δ' is defined as follows:

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 $p, q, r \in Q,\ a \in \Sigma$.

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

For results on suffix-closure and suffix-freeness, see Theorem 7 and Remark 6, 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 $Q' = Q \times Q \times Q$; $q_0' = [q_0, q_0, q_0]$; $F' = F \times (Q \setminus F) \times F$; and

$$\delta'([p, q, r], a) = \{[\delta(p, a), q, r],\ [\delta(p, a), \delta(q, a), r],\ [\delta(p, a), \delta(q, a), \delta(r, a)]\}, \quad \text{for all } p, q, r \in Q \text{ and } a \in \Sigma.$$

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 7 and Remark 6, 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 $(00)^*$), or it could fail because there is at least one, but only finitely many witnesses (for example, the language $10 + 0^*$).

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

Theorem 8. *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. The proof for each case is based on each NFA-ε construction given in Sections 2.1–2.4, respectively. Consider the NFA-ε 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-ε, so there is one additional complication: the cycle we find might be labeled completely by ε-transitions. To solve this, we use an idea suggested to us by Jack Zhao and Timothy Chan (personal communication). First, we find all the connected components of the transition graph of M', which can be done in linear time. Then, for each edge (p, q) labeled with something other than ε, which corresponds 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 ε. This technique also runs in linear time in the size of the NFA-ε [23]. □

2.5.1. 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 lack of the prefix-closure. Analogously, we define *almost suffix-*, *almost factor-*, and *almost subword-closed*.

Theorem 9. *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. The proof is analogous to the proof of Theorem 8 and is based on the NFA-ε construction in Theorem 3 and its variations. □

2.5.2. 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 lack of the prefix-freeness. Analogously, we define *almost suffix-*, *almost factor-*, and *almost subword-free*.

Theorem 10. *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. The proof is analogous to the proof of Theorem 8 and is based on the NFA-ε constructions in Remark 5 and its variations. □

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 11. *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. Theorem 1 proves the existence of an NFA-ε 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. □

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

Theorem 12. *There is a class of non-factor-convex regular languages L_n, accepted by DFA's with $O(n)$ states, such that the size of the minimal witness is $\Omega(n^3)$.*

The proof is postponed to Section 3.3 below.

Results analogous to Corollary 11 hold for prefix-, suffix-, and subword-convex languages. However, in some cases we can do better, as we show later.

3.1.1. Factor-closure

Theorem 3 gives us an $O(n^2)$ upper bound on the length of a witness to the lack of the factor-closure:

Corollary 13. *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

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

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

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

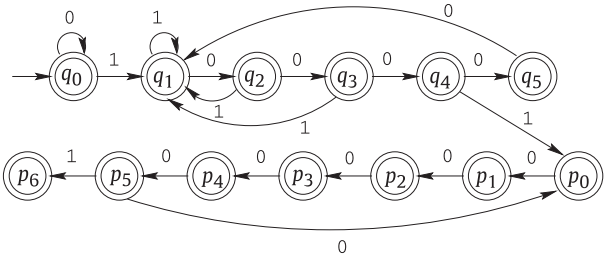

Fig. 1. Example of the construction in Theorem 14 for $n = 5$. All unspecified transitions go to a rejecting "dead state" q_6, not shown in the figure, that cycles on all inputs.

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

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

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

The DFA M has $2n + 4$ states. The language of M is denoted by the regular expression $L(M) = L_q + L_p$, where

$$L_q = \mathtt{0}^* + \mathtt{0}^*\mathtt{1}(\mathtt{1} + \mathtt{01} + \mathtt{0}^2\mathtt{1} + \cdots + \mathtt{0}^{n-3}\mathtt{1} + \mathtt{0}^n)^*(\varepsilon + \mathtt{0} + \mathtt{0}^2 + \cdots + \mathtt{0}^{n-2} + \mathtt{0}^{n-1}),$$
$$L_p = \mathtt{0}^*\mathtt{1}(\mathtt{1} + \mathtt{01} + \mathtt{0}^2\mathtt{1} + \cdots + \mathtt{0}^{n-3}\mathtt{1} + \mathtt{0}^n)^*\mathtt{0}^{n-2}\mathtt{1}(\mathtt{0}^{n+1})^*(\varepsilon + \mathtt{0} + \mathtt{0}^2 + \cdots + \mathtt{0}^n + \mathtt{0}^n\mathtt{1}).$$

For $n = 5$, M is illustrated in Fig. 1.
Then we have the following theorem:

Theorem 14. *For the DFA M above, let $L = L(M)$. For any witness (u, v) to the lack of factor-closure we have $|v| \ge n^2 + 2n$, and this bound is achievable.*

Proof. First, let us see that $L(M)$ is not factor-closed. Consider the words $u' = \mathtt{1}\mathtt{0}^{n^2+n-1}\mathtt{1}$ and $v' = \mathtt{1}\mathtt{0}^{n-2}u'$. Then it is easy to see that v' can be factored as $v' = \varepsilon \cdot \mathtt{1} \cdot \varepsilon \cdot \mathtt{0}^{n-2}\mathtt{1} \cdot (\mathtt{0}^{n+1})^{n-1} \cdot \mathtt{0}^n\mathtt{1}$, which shows that $v \in L_p$, and so $v' \in L(M)$. However, $z' = \mathtt{1}\mathtt{0}^{n^2+n-1} = (\mathtt{0}^n)^n\mathtt{0}^{n-1}$ takes M to state q_n, and $u' = z'\mathtt{1}$ takes M to the rejecting dead state q_{n+1}. Hence $u' \notin L(M)$. Thus we have a witness (u', v') demonstrating lack of factor closure. Note that $|v'| = n^2 + 2n$.

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

We now prove that the set of states along the rejecting path of u includes only states of type q. Suppose it includes both states of type q and type p. Since there is only one transition from a state of type q to a state of type p, and all transitions from a state of type p to a state of type q are to the rejecting state q_{n+1}, we have $u = u_1u_2$, where $\delta(q_0, u_1) = q_{n-1}$, and

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

Since u is a suffix of v, the last letter of v is also $\mathtt{1}$. So, by the construction of M, we have that $v = v_1v_2$, where $\delta(q_0, v_1) = q_{n-1}$, and

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

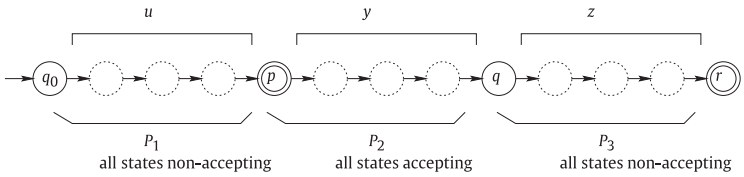

Fig. 2. The acceptance path for w.

It is obvious that $(\Sigma^* L_1) \cap (\Sigma^* L_2) = \emptyset$, which contradicts the equality $v_1 v_2 = v = wu = wu_1 u_2$. Therefore, the set of states along the rejecting path of u includes only states of type q.

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 $v \in \Sigma^* 1(0^{n+1})^* 0^n 1$. Thus, for some $k_u, k_v \geq 0$, one has $k_u n + n - 1 = k_v(n+1) + n$. One verifies that the smallest solution of this equation is $k_u = n$ and $k_v = n - 1$. Hence $|u| \geq k_u n + n - 1 + 2 = n^2 + n + 1$. Since q_{n-1} is the only state having a transition to a state of type p on input 1, and the shortest word that leads to state q_{n-1} is 10^{n-2}, we also have $|v| \geq 1 + n - 2 + n^2 + n + 1 = n^2 + 2n$. The witness (u', v') achieves this bound, and so the theorem is proved. □

3.1.2. Factor-freeness

From Remark 5, we get the following consequence:

Corollary 15. *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 15 is best possible, as the following theorem shows.

Theorem 16. *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 $L = 11(0^n)^+ 1 \cup 1(0^{n+1})^+ 1$. This language can be accepted by a DFA with $2n + 6$ states. The shortest witness to the lack of factor-freeness is $(10^{n(n+1)}1, 110^{n(n+1)}1)$, which has size $n^2 + n + 3$. □

3.2. Prefix-convexity

For prefix-convexity, we have the following theorem:

Theorem 17. *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 uyz, and let P_1 be the sequence of states from q_0 to p (not including p), P_2 be the sequence of states from p to q (not including q), and P_3 be the sequence of states from q to r (not including r); see Fig. 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.

Let r_i be the number of states in 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 the sets of states in 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 sets of states of P_3 are disjoint from those of P_2. So $r_1 + r_2 \leq n$ and $r_2 + r_3 \leq n$. It follows that $r_1 + r_2 + r_3 \leq 2n - r_2$. Since $r_2 \geq 1$, it follows that $|w| \leq 2n - 1$.

To see that $2n - 1$ is optimal, consider the DFA of n states accepting the unary language $L = 0^{n-1}(0^n)^*$. Then L is not prefix-convex, and the shortest witness is $(0^{n-1}, 0^n, 0^{2n-1})$. □

3.2.1. Prefix-closure

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

Theorem 18. *Let M be an n-state DFA, and suppose $L = L(M)$ is not prefix-closed. Then the minimal witness (v, w) showing that L is not prefix-closed has $|w| \leq n$, and this is best possible.*

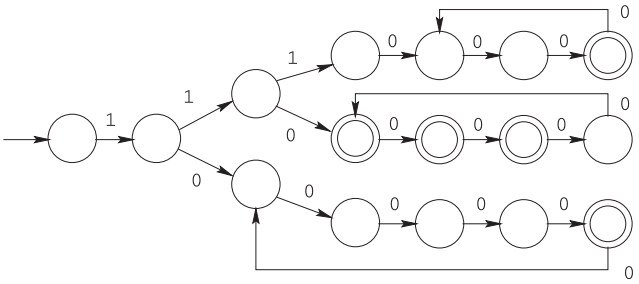

Fig. 3. Example of the construction in Theorem 20 for $n = 4$. All unspecified transitions go to a rejecting "dead state", not shown in the figure, that cycles on all inputs.

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 $L = (\mathtt{0}^n)^*$ shows. This language is not prefix-closed, can be accepted by a DFA with n states, and the smallest witness is $(\mathtt{0}, \mathtt{0}^n)$. □

3.2.2. Prefix-freeness

For the prefix-free property we have:

Theorem 19. *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 17. The bound is achieved by a unary DFA accepting $\mathtt{0}^{n-1}(\mathtt{0}^n)^*$. □

3.3. Suffix-convexity

For the suffix-convex property, the cubic upper bound implied by Corollary 11 is best possible, up to a constant multiplicative factor.

Theorem 20. *There is a class of non-suffix-convex regular languages L_n, accepted by DFA's with $O(n)$ states, such that the size of the minimal witnesses is $\Omega(n^3)$.*

Proof. Let $L = \mathtt{111}(\mathtt{0}^{n-1})^+ \cup \mathtt{11}(\mathtt{0} + \mathtt{00} + \cdots + \mathtt{0}^{n-1})(\mathtt{0}^n)^* \cup \mathtt{1}(\mathtt{0}^{n+1})^+$. Then L can be accepted by a DFA with $3n + 5$ states, as illustrated in Fig. 3.

Suppose (u, v, w) is a witness; then w cannot be a word of the form $\mathtt{10}^i$, because no proper suffix of such a word is in L. Also, w cannot be a word of the form $\mathtt{110}^i$, because the only proper suffix in L is $u = \mathtt{10}^i$. 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 $\mathtt{1110}^i$. The only proper suffixes of w in L are of the form $\mathtt{110}^i$ and $\mathtt{10}^i$. But we cannot have $u = \mathtt{110}^i$ because, if we did, there would be no v strictly between u and w in the suffix order. So it must be that $u = \mathtt{10}^i$. Then the only word in Σ^* strictly between u and w in the suffix order is $v = \mathtt{110}^i$, 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 $(\mathtt{10}^i, \mathtt{110}^i, \mathtt{1110}^i)$, where $i = \mathrm{lcm}(n - 1, n, n + 1) \geq (n - 1)n(n + 1)/2$. □

Now we consider the proof of Theorem 12. A similar technique can be used for non-factor-convex languages. This allows us to prove Theorem 12 in the same way we prove Theorem 20, except we use the language $L\mathtt{1}$ instead.

3.3.1. Suffix-closure

Obviously, a witness to the lack of suffix-closure is also a witness to the lack of factor-closure. So the proof of Theorem 14 shows that the bound $n^2 + 2n$ also holds for suffix-closed languages. However, Gill and Kou [11] showed that the bound $(n - 1)^2$ holds, and Veloso and Gill [13] improved this bound to $(n - 1)^2 - m(n - m) + 1$, where m is the number of accepting states of the DFA.

Ang and Brzozowski pointed out [1,2] 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.

Theorem 21. *Let M be a DFA of n states, and $L = L(M)$. Let v be one of the shortest words such that (u, v) is a witness to the lack of factor-closure for some word u. If $|v| > n$, then (u, v) is also a witness to the lack of suffix-closure.*

Proof. Suppose u is not a suffix of v. Write $v = v'a$ for $a \in \Sigma$. Then u is a factor of v'. So $v' \notin L$, for otherwise (u, v') is a shorter witness. Since $|v'| \geq n$, by the pumping lemma, we can write $v' = xyz$, where $xz \notin L$, $|xz| < |xyz|$. In addition, we have $xza \in L$, since $xyza = v'a = v \in L$. Then (xz, xza) is a witness to the lack of factor-closure and xza is shorter than v. This contradicts the fact that v is a shortest word of this kind. Therefore, u is a suffix of v and thus (u, v) is a witness to the lack of suffix-closure. □

3.3.2. Suffix-freeness

Theorem 22. *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 $L = \mathtt{11}(\mathtt{0}^n)^+ \cup \mathtt{1}(\mathtt{0}^{n+1})^+$. This language is accepted by a DFA with $2n + 5$ states. The shortest witness $(\mathtt{10}^{n(n+1)}, \mathtt{110}^{n(n+1)})$ to the lack of suffix-freeness has size $n^2 + n + 2$. □

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 with $n \leq m$. Let $q \in Q$ and $w \in L = L(M)$, and consider the state sequence

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

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 $\delta(q, w[1, i])$ with $j > i$, 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*.

We will first discuss the special cases of subword-closure and subword-freeness and postpone the discussion on subword-convexity to the end of this section.

3.4.1. 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 $v \trianglelefteq w$.

Theorem 23. *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 $w' \trianglelefteq w$. The theorem then follows.

Suppose that (v, w) is a witness, and $|w| = m \geq n+1$. Without loss of generality, we assume that (v, w) is minimal. Let $w = xyz$ be the canonical factorization of w with respect to the initial state, where $|xy| \leq n$, $|y| > 0$, and $|z| \geq m - n > 0$. Then, $xz \in L$. We will show that all the states M visits when reading w are accepting.

If there is a z' such that $z' \trianglelefteq z$ 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 $xz \trianglelefteq w$. Thus

$$z' \trianglelefteq z \text{ implies } xyz' \in L. \tag{1}$$

Consequently, we know that all the states that M visits reading z, including $\delta(q_0, xy)$, are accepting.

Since $v \trianglelefteq w = xyz$, we can write $v = v_x v_y v_z$ for some v_x, v_y, v_z such that $v_x \trianglelefteq x$, $v_y \trianglelefteq y$, and $v_z \trianglelefteq z$. Clearly, $v \trianglelefteq xyv_z$. 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 $xyv_z \trianglelefteq w$. Thus we deduce that the witness (v, w) has the form $(v_x v_y z, xyz)$.

If $y' \trianglelefteq y$ and $xy' \notin L$, then (xy', xy) is a witness with $|xy| < |w|$ and $xy \trianglelefteq w$. Thus

$$y' \trianglelefteq y \text{ implies } xy' \in L. \tag{2}$$

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

$$x' \trianglelefteq x \text{ implies } x' \in L. \tag{3}$$

From (1)–(3), we conclude that all the states M visits when reading w are accepting. We know that the states in the sequence

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

are all distinct, in view of the canonical decomposition of w. Also, the states in the sequence

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

are all accepting and distinct; otherwise, v would not be the shortest for w such that (v, w) is a witness.

We now claim that no state can be in both S and S'. First, 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 $xz \trianglelefteq w$, since $w[1, i] = x[1, i]$, and $x[1, i]z[k + 1, |z|] \trianglelefteq xz$. 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 $xyz[k+1, |z|] \trianglelefteq w$, since $xy[1, j]z[k+1, |z|] \trianglelefteq xyz[k+1, |z|]$, and $xyz[k+1, |z|] \in L$ by (1).

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

Corollary 24. *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 Theorem 23, it has a witness (v, w) with $|w| \leq n$. This is the best possible bound for $n \geq 2$, since the language $(\mathtt{0}^n)^*(\varepsilon + \mathtt{0} + \cdots + \mathtt{0}^{n-2})$, accepted by a DFA with n states, has a minimal witness $(\mathtt{0}^{n-1}, \mathtt{0}^n)$. □

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

3.4.2. Subword-freeness

Theorem 25. *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 $w' \trianglelefteq w$.*

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 $w' \trianglelefteq w$. The theorem 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 $xz \trianglelefteq w$. □

Corollary 26. *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 Theorem 25, it has a witness (v, w) with $|w| \leq 2n - 1$. This is the best possible bound for $n \geq 2$, since the language $\mathtt{0}^{n-1}(\mathtt{0}^n)^*$, accepted by a DFA with n states, has a minimal witness $(\mathtt{0}^{n-1}, \mathtt{0}^{2n-1})$. □

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

3.4.3. Subword-convexity in general cases

Before discussing witnesses to the lack of subword-convexity, we require a result that is a consequence of the pumping lemma.

Lemma 27. *Let M be a DFA with $n \geq 2$ states, and let $w \in L = L(M)$ satisfy $|w| \geq 2n$. Then there exists a factorization $w = x_1y_1x_2y_2 \cdots x_ky_kz_k$ such that $k \geq 2$, $x_1y_1^*x_2y_2^* \cdots x_ky_k^*z_k \subseteq L$, $|y_i| > 0$, for $i = 1, \ldots, k$, and $|x_2 \cdots x_kz_k| < n$.*

Proof. Suppose $w \in L$ and $|w| \geq 2n$. Let $w = x_1y_1z_1$ be the canonical factorization of w with respect to the initial state of M; hence $|x_1y_1| \leq n$, $|y_1| > 0$, and $|z_1| \geq n \geq 2$. Then $x_1z_1 \in L$. Consider the following sequence of states:

$$p_0 = \delta(q_0, x_1y_1),\ \ p_1 = \delta(q_0, x_1y_1z_1[1, 1]), \ldots, p_{|z_1|} = \delta(q_0, x_1y_1z_1).$$

Since $|z_1| \geq n$, there must be at least one pair (p_i, p_l), $l > i$, of states such that $p_i = p_l$. If $i = 0$, let j be the greatest index such that $p_0 = p_j$, let $x_2 = \varepsilon$, let $y_2 = z_1[1, j]$, and let $z_2 = z_1[j + 1, |z_1|]$. If $i > 0$, then 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 $z_2 = z_1[j + 1, |z_1|]$.

Observe that in the canonical factorization we choose the first state that appears at least twice and the *first* occurrence of that state. In the factorization of z_1 above, we also choose the first state that appears at least twice, but then we take the *last* occurrence of that state.

If the sequence $\delta(q_0, x_1y_1x_2y_2), \delta(q_0, x_1y_1x_2y_2z_2[1,1]), \ldots, \delta(q_0, x_1y_1x_2y_2z_2)$ 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

$$w = x_1y_1x_2y_2\cdots x_ky_kz_k,$$

where $x_1y_1^*x_2y_2^*\cdots x_ky_k^*z_k \subseteq L$, $|y_i| > 0$, for $i = 1, \ldots, k$, and $k \geq 2$. We also have $|x_2\cdots x_kz_k| < n$; otherwise, there would be repeated states. By construction, there are no repeated states in $x_1x_2\cdots x_kz_k$, and the factorization satisfies all the requirements of the lemma. □

Theorem 28. *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 $w' \trianglelefteq w$, 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 $w' \trianglelefteq w$. The theorem then follows.

We assume without loss of generality that w is a minimal witness. First, consider the witness (u, v) to the lack of subword-closure of the language $\Sigma^* \setminus L$. By Theorem 23, there exists a witness (u', v') to the lack of subword-closure of $\Sigma^* \setminus L$ such that $v' \trianglelefteq v$ and $|v'| \leq n$. Thus there must be a witness (u, v, w) to the lack of subword-convexity such that $|v| \leq n$.

Suppose that (u, v, w) is a witness, and $|w| \geq 3n - 1$. By Lemma 27, w has the factorization $w = x_1y_1x_2y_2\cdots x_ky_kz_k$. For any $y_2' \trianglelefteq y_2, \cdots, y_k' \trianglelefteq y_k$, we have $x_1y_1x_2y_2'\cdots x_ky_k'z_k \in L$. Otherwise, the triple

$$(x_1x_2\cdots x_kz_k,\ x_1x_2y_2'\cdots x_ky_k'z_k,\ x_1x_2y_2\cdots x_ky_kz_k)$$

is a witness with $|x_1x_2y_2\cdots x_ky_kz_k| < |w|$, and $x_1x_2y_2\cdots x_ky_kz_k \trianglelefteq w$.

Since $v \trianglelefteq w$, 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 $v_t \trianglelefteq t$ for $t = x_1, y_1, \ldots, x_k, y_k, z_k$. If there is a y_i with $i \geq 2$, such that $v_{y_i} = \varepsilon$, then we can replace that y_i by ε in w and obtain a smaller witness. Hence each v_{y_i} must be nonempty. By the same argument, if $v_{y_i} \neq y_i$, for $i \geq 2$, then we can replace that y_i by v_{y_i} in w, yielding a smaller witness. Therefore $y_i = v_{y_i}$ for $i = 2, \ldots, k$. We claim that $|y_2\cdots y_k| < |v|$; otherwise $v = v_{y_2}\cdots v_{y_k} = y_2\cdots y_k$ and $(u, v, x_1x_2y_2\cdots x_ky_kz_k)$ is a witness with $|x_1x_2y_2\cdots x_ky_kz_k| < |w|$. Thus $|y_2\cdots y_k| < |v| \leq n$, and we conclude that

$$|w| = |x_1y_1| + |x_2\cdots x_kz_k| + |y_2\cdots y_k| \leq n + (n-1) + (n-1) = 3n - 2. \quad \square$$

Corollary 29. *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 $0^{n-1}(0^n)^*$ is accepted by a DFA with n states and has a minimal witness $(0^{n-1}, 0^n, 0^{2n-1})$, 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 the following result:

Theorem 30. *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) [24]: given a directed acyclic graph $G = (V, E)$ on the vertices $\{q_1, q_2, \ldots, q_n\}$, 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 $e_1, e_2, \ldots, e_m$; then there is a path that uses no edges into q_1 or out of q_n. In M this corresponds to the word $\#e_1e_2\cdots e_m\$$; 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 $\#e_1e_2\cdots e_m\$$. Since M accepts #, $L(M)$ fails to be prefix-free if and only it accepts a word of the form $\#e_1e_2\cdots e_m\$$. But then $e_1, e_2, \ldots, e_m$ 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 2.1.3 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 problems. □

5. Languages specified by other means

Now 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 [25]:

Lemma 31. *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 Δ 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 does not accept x, and $L(r_x) = \Delta^* \setminus \{w\}$ 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 $L(M_x) = \Delta^* \setminus \{w\}$ for some nonempty w (depending on x) otherwise.*

Theorem 32. *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. Similarly, deciding whether $\overline{L}$ is prefix-closed (respectively, suffix-, factor-, subword-closed) is also 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 2 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 11, 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.

The problem is PSPACE-hard: Since Δ^* is factor-convex and $\Delta^* \setminus \{w\}$ is not if $w \neq \varepsilon$, we could use an algorithm for the factor-convex problem to solve decidability for polynomial-space bounded Turing machines. □

For the prefix-, suffix-, and factor-closed properties, PSPACE-hardness was already proved by Hunt and Rosenkrantz [26, Theorem 3.4]; also see [14,15].

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. [22] through a different approach.

Theorem 33. *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). Hence it costs $O(n^2+t^2)$ to check whether $L(M'') = \emptyset$ using depth-first search.

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

The construction for subword-freeness is slightly more involved. Create M' by making two copies of M. Add a transition from each state q to its copy q' on each letter of Σ, and add transitions from each copy q' to itself on each letter of Σ. The final states of M' are the final states in the part corresponding to the copied states. Formally, $M' = (Q \cup Q', \Sigma, \delta, q_0, F')$ 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. □

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

Proof. In Ellul et al. [27, §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. □

For prefix-freeness, we have the following theorem:

Theorem 35. *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 22. □

For the lack of subword-freeness, we cannot improve the bound we obtained for DFA's in Corollary 26, 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, closed, and converse-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 [28, 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 [9, Theorem 9.5, p. 581].

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

Theorem 36. *There is an algorithm that, given a context-free grammar G, will decide if $L(G)$ is subword-free.*

Proof. It is well known that, if $L = L(G)$ is infinite, then L is not subword-free [4,7,9]. We can test if $L(G)$ is infinite by a well-known result [28, Theorem 6.6, p. 137]. Otherwise, if $L(G)$ is finite, we can enumerate all its words by a bottom-up examination of the grammar, and test each word for the subword-free property. □

We can also consider the shortest witness to the lack of the subword-freeness.

Theorem 37. *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 lack of the subword-freeness 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 $\leq 2^{n-1}$. Also, 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. □

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	$2n-1$

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 upper bound for subword-convexity; we do not know whether it is achievable.

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, and also for their improvements to several proofs.

References

[1] T. Ang, J. Brzozowski, Continuous languages, in: E. Csuhaj-Varjú, Z. Ésik (Eds.), Proceedings of the 12th International Conference on Automata and Formal Languages,, Computer and Automation Research Institute, Hungarian Academy of Sciences, 2008, pp. 74–85.
[2] T. Ang, J. Brzozowski, Languages convex with respect to binary relations, and their closure properties, Acta Cybernet. 19 (2009) 445–464.
[3] G. Thierrin, Convex languages, in: M. Nivat (Ed.), Automata, Languages, and Programming, North-Holland, 1973, pp. 481–492.
[4] L.H. Haines, On free monoids partially ordered by embedding, J. Combin. Theory 6 (1) (1969) 94–98.
[5] H. Jürgensen, S.S. Yu, Relations on free monoids, their independent sets, and codes, Internat. J. Comput. Math. 40 (1991) 17–46.
[6] H.J. Shyr, G. Thierrin, Hypercodes, Inform. and Control 24 (1974) 45–54.
[7] H.J. Shyr, Free Monoids and Languages, Hon Min Book Co., Taiwan, 2001.
[8] J. Berstel, D. Perrin, C. Reutenauer, Codes and Automata, Cambridge University Press, Cambridge, 2010.
[9] H. Jürgensen, S. Konstantinidis, Codes, in: G. Rozenberg, A. Salomaa (Eds.), Handbook of Formal Languages, vol. 1, Springer-Verlag, 1997, pp. 511–607.
[10] H. Jürgensen, L. Kari, G. Thierrin, Morphisms preserving densities, Internat. J. Comput. Math. 78 (2001) 165–189.
[11] A. Gill, L. Kou, Multiple-entry finite automata, J. Comput. System Sci. 9 (1974) 1–19.
[12] Z. Galil, J. Simon, A note on multiple-entry finite automata, J. Comput. System Sci. 12 (1976) 350–351.
[13] P.A.S. Veloso, A. Gill, Some remarks on multiple-entry finite automata, J. Comput. System Sci. 18 (1979) 304–306.
[14] J.-Y. Kao, N. Rampersad, J. Shallit, On NFAs where all States are Final, Initial, or Both, preprint, 2009. Available from: <http://arxiv.org/abs/0808.2417>.
[15] J.-Y. Kao, N. Rampersad, J. Shallit, On NFAs where all states are final, initial, or both, Theoret. Comput. Sci. 410 (47–49) (2009) 5010–5021. <http://dx.doi.org/10.1016/j.tcs.2009.07.049> .
[16] A. de Luca, S. Varricchio, Some combinatorial properties of factorial languages, in: R. Capocelli (Ed.), Sequences, Springer, 1990, pp. 258–266.
[17] A. Paz, B. Peleg, Ultimate-definite and symmetric-definite events and automata, J. ACM 12 (3) (1965) 399–410. <http://doi.acm.org/10.1145/321281.321292>.
[18] J.A. Brzozowski, J. Shallit, Z. Xu, Decision Problems for Convex Languages, preprint, 2008. Available from: <http://arxiv.org/abs/0808.1928>.
[19] J. Brzozowski, J. Shallit, Z. Xu, Decision problems for convex languages, in: A.H. Dediu, A.M. Ionescu, C. Martín-Vide (Eds.), LATA 2009, Lecture Notes in Computer Science, vol. 5457, Springer, 2009, pp. 247–258.
[20] T.H. Cormen, C.E. Leiserson, R.L. Rivest, Introduction to Algorithms, MIT Press, 2009.
[21] M.-P. Béal, M. Crochemore, F. Mignosi, A. Restivo, M. Sciortino, Computing forbidden words of regular languages, Fund. Inform. 56 (2003) 121–135.
[22] Y.-S. Han, Y. Wang, D. Wood, Infix-free regular expressions and languages, Internat. J. Found. Comput. Sci. 17 (2006) 379–393.
[23] R. Tarjan, Depth-first search and linear graph algorithms, SIAM J. Comput. 1 (1972) 146–160.
[24] N.D. Jones, Space-bounded reducibility among combinatorial problems, J. Comput. System Sci. 11 (1975) 68–85.
[25] A. Aho, J. Hopcroft, J. Ullman, The Design and Analysis of Computer Algorithms, Addison-Wesley, 1974.
[26] H.B. Hunt III, D.J. Rosenkrantz, Computational parallels between the regular and context-free languages, SIAM J. Comput. 7 (1978) 99–114.
[27] K. Ellul, B. Krawetz, J. Shallit, M.-W. Wang, Regular expressions: new results and open problems, J. Autom. Lang. Comb. 10 (2005) 407–437.
[28] J.E. Hopcroft, J.D. Ullman, Introduction to Automata Theory, Languages, and Computation, Addison-Wesley, 1979.