Acta Cybernetica **23** (2017) 9–41.

Complexity of Right-Ideal, Prefix-Closed, and Prefix-Free Regular Languages*

Janusz A. Brzozowski[a] and Corwin Sinnamon[a]

Abstract

A language L over an alphabet Σ is prefix-convex if, for any words $x, y, z \in \Sigma^*$, whenever x and xyz are in L, then so is xy. Prefix-convex languages include right-ideal, prefix-closed, and prefix-free languages as special cases. We examine complexity properties of these special prefix-convex languages. In particular, we study the quotient/state complexity of boolean operations, product (concatenation), star, and reversal, the size of the syntactic semigroup, and the quotient complexity of atoms. For binary operations we use arguments with different alphabets when appropriate; this leads to higher tight upper bounds than those obtained with equal alphabets. We exhibit right-ideal, prefix-closed, and prefix-free languages that meet the complexity bounds for all the measures listed above.

Keywords: atoms, complexity of operations, prefix-closed, prefix-convex, prefix-free, quotient complexity, regular languages, right ideals, state complexity, syntactic semigroup, unrestricted alphabets

I have known Zoltán Ésik for about 30 years. We met at many scientific conferences, seven of them in Hungary. In 2000 I invited Zoltán to spend a month in Waterloo so that we could work on a problem in algebra with which I was struggling. I thought that the problem was purely of theoretical interest, but it turned out that the algebra we discovered was applicable to the detection of hazards in logic circuits. Zoltán also helped me with two other algebraic problems; he was always ready to give advice, and was modest about taking credit for his contributions. As the years went by we became good friends. In June 2015 my wife and I had the honour of celebrating his 64th birthday at our house. His passing was a great shock to me and I greatly miss his friendship.

Janusz Brzozowski

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

[a] David R. Cheriton School of Computer Science, University of Waterloo,Waterloo, ON, Canada N2L 3G1, E-mail: `brzozo@uwaterloo.ca` `sinncore@gmail.com`

DOI: 10.14232/actacyb.23.1.2017.3

1 Motivation

For words w, x, y over an alphabet Σ, if $w = xy$, then x is a *prefix* of w. A language $L \subseteq \Sigma^*$ is *prefix-convex* [1, 28] if, whenever x and xyz are in L, then xy is also in L. The class of prefix-convex languages includes three well-known subclasses: right-ideal, prefix-closed, and prefix-free languages; we study complexity properties of these languages.

A language L is a *right ideal* if it is non-empty and satisfies the equation $L = L\Sigma^*$. Right ideals play a role in pattern matching: If one is searching for all words beginning with words in some language L in a given text (a word over Σ^*), then one is looking for words in $L\Sigma^*$. Right ideals also constitute a basic concept in semigroup theory.

A language L is *prefix-closed* if, whenever w is in L and x is a prefix of w, then x is also in L. The complement of every right ideal is a prefix-closed language. The set of allowed input sequences to any digital system is a prefix-closed language.

A language L is *prefix-free* if no word in L is a prefix of another word in L. Prefix-free languages (other than $\{\varepsilon\}$, where ε is the empty word) are prefix codes. They play an important role in coding theory, and have many applications [3].

The *alphabet of a regular language* L is Σ (or L *is a language over* Σ) if $L \subseteq \Sigma^*$ and every letter of Σ appears in a word of L. The *(left) quotient* of L by a word $w \in \Sigma^*$ is $w^{-1}L = \{x \mid wx \in L\}$. A language is regular if and only if it has a finite number of distinct quotients. So the number of quotients of L is a natural measure of complexity for L; it is called the *quotient complexity* [4] of L and is denoted it by $\kappa(L)$. An equivalent concept is the *state complexity* [29] of L, which is the number of states in a complete minimal deterministic finite automaton (DFA) with alphabet Σ recognizing L.

If L_n is a regular language of quotient complexity n, and $\circ$ is a unary operation, then the *quotient/state complexity of* $\circ$ is the maximal value of $\kappa(L_n^\circ)$, expressed as a function of n, as L_n ranges over all regular languages of complexity n. If L'_m and L_n are regular languages of quotient complexities m and n respectively, and $\circ$ is a binary operation, then the *quotient/state complexity of* $\circ$ is the maximal value of $\kappa(L'_m \circ L_n)$, expressed as a function of m and n, as L'_m and L_n range over all regular languages of complexities m and n, respectively. The quotient/state complexity of an operation gives a worst-case lower bound on the time and space complexities of the operation, and has been studied extensively [4, 5, 29]; we refer to quotient/state complexity simply as *complexity.*

In all the past literature on binary operations it has always been assumed that the alphabets of the two operands are restricted to be the same. However, it has been shown recently [6, 14] that this is an unnecessary restriction: larger complexity bounds can be reached in some cases if the alphabets differ. In the present paper we examine both *restricted complexity* of binary operations, where the alphabets must be the same, and *unrestricted complexity*, where they may differ.

To find the complexity of a unary operation one first finds an upper bound on this complexity, and then exhibits languages that meet this bound. Since we require a language L_n for each $n \geq k$, we need a sequence $(L_k, L_{k+1}, \dots)$; here k is usually a small integer because the bound may not hold for a few small values of n. We call such a sequence a *stream* of languages. Usually the languages in a stream have the same basic structure and differ only in the parameter n. For example, $((a^n)^* \mid n \geq 2)$ is a stream. For a binary operation we require two streams.

While the complexity of languages is a useful measure, it is not entirely satisfactory. Two languages may have the same complexity n but the syntactic semigroup [26] of one may have $n-1$ elements, while that of the other has n^n elements [18]. For this reason, the size of the syntactic semigroup of a language – which is the same as the size of the transition semigroup of a minimal DFA accepting the language [26] – has been added as another complexity measure. Secondly, *star-free* languages meet the complexity bounds of regular languages for all operations except reversal, which only reaches the bound $2^n - 1$ instead of 2^n [13]. While regular languages are the smallest class containing the finite languages and closed under boolean operations, product and star, star-free languages are the smallest class containing the finite languages and closed only under boolean operations and product. In view of the results in [13], quotient/state complexity does not distinguish between these two classes.

The complexities of the atoms of a regular language have been proposed as an additional measure [5]. Atoms are defined by the following left congruence: two words x and y are equivalent if $ux \in L$ if and only if $uy \in L$ for all $u \in \Sigma^*$. Thus x and y are equivalent if $x \in u^{-1}L$ if and only if $y \in u^{-1}L$. An equivalence class of this relation is an *atom* of L [17, 21]. Thus an atom is a non-empty intersection of complemented and uncomplemented quotients of L. If $K_0, \dots, K_{n-1}$ are the quotients of L, and $S \subseteq Q_n = \{0, \dots, n-1\}$, then atom A_S is the intersection of quotients with subscripts in S and complemented quotients with subscripts in $Q_n \setminus S$. For more information about atoms see [16, 17, 21].

There exists a stream $(L_3, L_4, \dots)$ of regular languages $L_n(a, b, c)$ that meets the restricted complexity bounds for all boolean operations, product (concatenation), star, and reversal, and also has the largest syntactic semigroup and most complex atoms [5]. This stream modified by the addition of an input d that performs the identity transformation also meets the unrestricted bounds for product and boolean operations [6, 14]; such a stream is called *most complex*. Most complex streams are useful when one designs a system dealing with regular languages and finite automata. If one would like to know the maximal sizes of automata the system can handle, one can use the one most complex stream to test all the operations.

2 Contributions

We first present a most complex regular language stream similar to that of [5], but one that is better suited for prefix-convex languages. We then exhibit most complex language streams for right-ideal, prefix-closed, and prefix-free languages. More specifically, our contributions are as follows:

1. We generalize the concept of permutational dialect defined in [5, 9] by allowing letters of an alphabet to be mapped to letters from a different alphabet.

2. For regular languages we prove that there exists a most complex language stream $(L_n(a,b,c) \mid n \geq 3)$. The following results are new:
   - $L'_m(a,b)L_n(a,-,b)$ and $L'_m(a,b)L_n(a,c,b)$ meet the known bounds $(m-1)2^n + 2^{n-1}$ and $m2^n + 2^{n-1}$ for restricted and unrestricted products, respectively.
   - For the unrestricted case the following hold:
     - $L'_m(a,b,c) \circ L_n(b,a,d)$ meets the known bound $(m+1)(n+1)$ when $\circ \in \{\cup, \oplus\}$, where $\oplus$ is symmetric difference.
     - $L'_m(a,b,c) \setminus L_n(b,a)$ meets the known bound $mn + m$.
3. For right-ideal languages we prove that there exists a most complex language stream $(L_n(a,b,c,d) \mid n \geq 4)$. The following results are new:
   - $L'_m(a,-,c,d)L_n(a,-,c,d)$ meets the known bound $m + 2^{n-2}$ for restricted product, and $L'_m(a,-,c,d)L_n(b,-,c,d)$ meets the bound $m + 2^{n-1} + 2^{n-2} + 1$ for unrestricted product.
   - For the restricted case the known bounds mn if $\circ \in \{\cap, \oplus\}$, $mn-(m-1)$ if $\circ = \setminus$, and $mn-(m+n-2)$ if $\circ = \cup$ are all met by $L'_m(a,-,-,d) \circ L_n(-,-,d,a)$.
   - For the unrestricted case the bounds are the same as for regular languages and they are met by $L'_m(a,-,c,d) \circ L_n(b,-,d,a)$ if $\circ \in \{\cup, \oplus\}$, $L'_m(a,-,c,d) \setminus L_n(-,-,d,a)$, and $L'_m(a,-,-,d) \cap L_n(-,-,d,a)$.
4. For prefix-closed languages we prove that there exists a most complex language stream $(L_n(a,b,c,d) \mid n \geq 4)$. Here restricted and unrestricted cases coincide. The following results are new:
   - $L'_m(a,b,c,d)L_n(a,d,b,c)$ meets the known bound $(m+1)2^{n-2}$.
   - The known bounds mn if $\circ \in \{\cup, \oplus\}$, $mn-(m-1)$ if $\circ = \setminus$, and $mn-(m+n-2)$ if $\circ = \cup$ are met by $L'_m(a,b,-,d) \circ L_n(b,a,-,d)$.
5. For prefix-free languages we prove that there exists a most complex language stream $(L_n(a,b,c,d,e_0,\dots,e_{n-3}) \mid n \geq 4)$; restricted and unrestricted cases coincide. The following results are new:
   - At least $n+2$ inputs are required for a most complex prefix-free witness.
   - At least $n+1$ inputs are necessary to reach the known bound n^{n-2} for the size of the syntactic semigroup.
   - We derive upper bounds for the complexity of atoms of prefix-free languages, and prove that the atoms of the language $L_n(a,b,c,-,e_0)$ meet these bounds.
   - $L'_m(a,b,c,d)L_n(a,d,b,c)$ meets the known bound $(m+1)2^{n-2}$.
   - The known bounds $mn-2$ if $\circ \in \{\cup, \oplus\}$, $mn-(m+2n-4)$ if $\circ = \setminus$, and $mn-2(m+n-3)$ if $\circ = \cap$ are met by $L'_m(a,b,-,-,e_0,e_{m-3}) \circ L_n(b,a,-,-,e_0,e_{m-3})$.

3 Finite Automata, Transformations, Semigroups

A *deterministic finite automaton (DFA)* is a quintuple $\mathcal{D} = (Q, \Sigma, \delta, q_0, F)$, where Q is a finite non-empty set of *states*, Σ is a finite non-empty *alphabet*, $\delta\colon Q \times \Sigma \to Q$ is the *transition function*, $q_0 \in Q$ is the *initial* state, and $F \subseteq Q$ is the set of *final* states. We extend δ to a function $\delta\colon Q \times \Sigma^* \to Q$ as usual. A DFA $\mathcal{D}$ *accepts* a word $w \in \Sigma^*$ if $\delta(q_0, w) \in F$. The language accepted by $\mathcal{D}$ is denoted by $L(\mathcal{D})$. If q is a state of $\mathcal{D}$, then the language L^q of q is the language accepted by the DFA $(Q, \Sigma, \delta, q, F)$. A state is *empty* or *dead* or *a sink* if its language is empty. Two states p and q of $\mathcal{D}$ are *equivalent* if $L^p = L^q$; otherwise they are *distinguishable.* A state q is *reachable* if there exists $w \in \Sigma^*$ such that $\delta(q_0, w) = q$. A DFA is *minimal* if all of its states are reachable and no two states are equivalent. Usually DFAs are used to establish upper bounds on the complexity of operations and also as witnesses that meet these bounds.

A *nondeterministic finite automaton (NFA)* is a quintuple $\mathcal{D} = (Q, \Sigma, \delta, I, F)$, where Q, Σ and F are as in a DFA, $\delta\colon Q \times \Sigma \to 2^Q$, and $I \subseteq Q$ is the *set of initial states.* An *ε-NFA* is an NFA in which transitions under the empty word ε are also permitted.

Without loss of generality we use $Q_n = \{0, \dots, n-1\}$ as the set of states of every DFA with n states. A *transformation* of Q_n is a mapping $t\colon Q_n \to Q_n$. The *image* of $q \in Q_n$ under t is denoted by qt. In any DFA, each letter $a \in \Sigma$ induces a transformation δ_a of the set Q_n defined by $q\delta_a = \delta(q, a)$; we denote this by $a\colon \delta_a$. By a slight abuse of notation we use the letter a to denote the transformation it induces; thus we write qa instead of $q\delta_a$. We extend the notation to sets of states: if $P \subseteq Q_n$, then $Pa = \{pa \mid p \in P\}$. We also write $P \xrightarrow{a} Pa$ to mean that the image of P under a is Pa. Let $\mathcal{T}_{Q_n}$ be the set of all n^n transformations of Q_n; then $\mathcal{T}_{Q_n}$ is a monoid under composition.

For $k \ge 2$, a transformation (permutation) t of a set $P = \{q_0, q_1, \dots, q_{k-1}\} \subseteq Q_n$ is a *k-cycle* if $q_0t = q_1, q_1t = q_2, \dots, q_{k-2}t = q_{k-1}, q_{k-1}t = q_0$. This k-cycle is denoted by the transformation $(q_0, q_1, \dots, q_{k-1})$ of Q_n, which acts as the identity on the states outside the cycle. A 2-cycle (q_0, q_1) is called a *transposition.* A transformation that sends all the states of P to q and acts as the identity on the remaining states is denoted by $(P \to q)$. If $P = \{p\}$ we write $(p \to q)$ for $(\{p\} \to q)$. The identity transformation is denoted by $\mathbb{1}$. The notation $(_i^j\ q \to q+1)$ denotes a transformation that sends q to $q+1$ for $i \le q \le j$ and is the identity for the remaining states, and $(_i^j\ q \to q-1)$ is defined similarly.

Let $\mathcal{D} = (Q_n, \Sigma, \delta, q_0, F)$ be a DFA. For each word $w \in \Sigma^*$, the transition function induces a transformation δ_w of Q_n by w: for all $q \in Q_n$, $q\delta_w = \delta(q, w)$. The set $T_{\mathcal{D}}$ of all such transformations by non-empty words forms a semigroup of transformations called the *transition semigroup* of $\mathcal{D}$ [26]. We can use a set $\{\delta_a \mid a \in \Sigma\}$ of transformations to define δ, and so the DFA $\mathcal{D}$.

The *Myhill congruence* [25] $\approx_L$ of a language $L \subseteq \Sigma^*$ is defined on Σ^+ as follows:

$$\text{For } x, y \in \Sigma^+,\ x \approx_L y \text{ if and only if } wxz \in L \Leftrightarrow wyz \in L \text{ for all } w, z \in \Sigma^*.$$

This congruence is also known as the *syntactic congruence* of L. The quotient set

$\Sigma^+/{\approx_L}$ of equivalence classes of the relation $\approx_L$ is a semigroup called the *syntactic semigroup* of L. If $\mathcal{D}$ is a minimal DFA of L, then $T_{\mathcal{D}}$ is isomorphic to the syntactic semigroup T_L of L [26], and we represent elements of T_L by transformations in $T_{\mathcal{D}}$. The size of the syntactic semigroup has been used as a measure of complexity for regular languages [5, 18, 20, 24].

Recall that binary operations require two language streams to determine the complexity of the operation. Sometimes the same stream can be used for both operands, and it has been shown in [5, 6] that for all common binary operations on regular languages the second stream can be a "dialect" of the first, that is, it can "differ only slightly" from the first and all the bounds can still be met. Let $\Sigma = \{a_1, \dots, a_k\}$ be an alphabet ordered as shown; if $L \subseteq \Sigma^*$, we denote it by $L(a_1, \dots, a_k)$ to stress its dependence on Σ. A *dialect* of L is a related language obtained by replacing or deleting letters of Σ in the words of L. More precisely, for an alphabet Σ' and a partial map $\pi\colon \Sigma \mapsto \Sigma'$, we obtain a dialect of L by replacing each letter $a \in \Sigma$ by $\pi(a)$ in every word of L, or deleting the word entirely if $\pi(a)$ is undefined. We write $L(\pi(a_1), \dots, \pi(a_k))$ to denote the dialect of $L(a_1, \dots, a_k)$ given by π, and we denote undefined values of π by "$-$". For example, if $L(a, b, c) = \{a, ab, ac\}$ then its dialect $L(b, -, d)$ is the language $\{b, bd\}$. Undefined values for letters at the end of the alphabet are omitted; thus, for example, if $\Sigma = \{a, b, c, d, e\}$, $\pi(a) = b$, $\pi(b) = a$, $\pi(c) = c$ and $\pi(d) = \pi(e) = -$, we write $L(b, a, c)$ for $L(b, a, c, -, -)$.

The language stream that meets all the complexity bounds is referred to as the *master* language stream. Every master language stream we present here uses the smallest possible alphabet sufficient to meet all the bounds. Individual bounds are frequently met by dialects on reduced alphabets, and we prefer to use the smallest alphabet possible for each bound. For binary operations, we try to minimize the size of the combined alphabet of the two dialects.

As each letter induces a transformation on the states of a DFA (or equivalently, the quotients of a language) we count the number of distinct transformations induced by letters of the alphabet. In any language this number is at most the size of the alphabet, but there may be multiple letters which induce the same transformation; this does not occur in this paper as no language has a repeated transformation. For binary operations on two dialects of the same master language, we count the number of distinct transformations of the master language present in either dialect. For example, suppose $L(a, b, c, -)$ and $L(a, -, b, c)$ are two dialects of a language $L(a, b, c, d)$, which we assume has four distinct transformations. Each dialect has three letters and three distinct transformations, and between them they have three letters and four distinct transformations.

Although a given complexity bound may be met by many dialects of the master language, we favour dialects, or pairs of dialects, that use small alphabets and few distinct transformations. In many cases the dialects we present are minimal in these respects, though we do not always prove this.

4 A Most Complex Regular Stream

We now define a DFA stream that we use as a basic component. It is similar to the stream defined in [5] for the case of equal alphabets, except that there the transformation induced by c is $(n-1 \to 0)$. It is also similar to the DFA of [6], except that there the transformation induced by c is $(n-1 \to 0)$ and an additional input d inducing the identity transformation is used.

Definition 1. *For $n \geq 3$, let $\mathcal{D}_n = \mathcal{D}_n(a,b,c) = (Q_n, \Sigma, \delta_n, 0, \{n-1\})$, where $\Sigma = \{a,b,c\}$, and δ_n is defined by the transformations $a\colon (0,\dots,n-1)$, $b\colon (0,1)$, and $c\colon (1 \to 0)$. Let $L_n = L_n(a,b,c)$ be the language accepted by $\mathcal{D}_n$. The structure of $\mathcal{D}_n(a,b,c)$ is shown in Figure 1.*

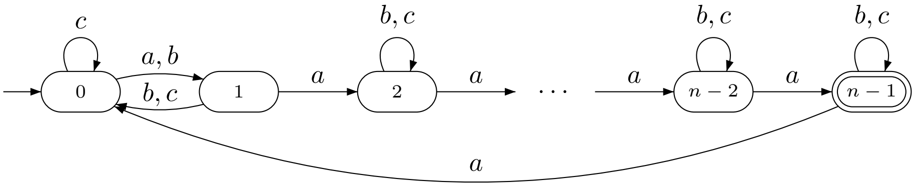

Figure 1: Minimal DFA of a most complex regular language.

Theorem 1 (Most Complex Regular Languages)**.** *For each $n \geq 3$, the DFA of Definition 1 is minimal and its language $L_n(a,b,c)$ has complexity n. The stream $(L_m(a,b,c) \mid m \geq 3)$ with some dialect streams is most complex in the class of regular languages. In particular, it meets all the complexity bounds below. At least three letters are required in any witness meeting all these bounds and a total of four distinct letters is required for any two witnesses for unrestricted union and symmetric difference. In several cases the bounds can be met with a smaller alphabet as shown below.*

1. *The syntactic semigroup of $L_n(a,b,c)$ has cardinality n^n.*
2. *Each quotient of $L_n(a)$ has complexity n.*
3. *The reverse of $L_n(a,b,c)$ has complexity 2^n, and $L_n(a,b,c)$ has 2^n atoms.*
4. *Each atom A_S of $L_n(a,b,c)$ has maximal complexity:*

$$\kappa(A_S) = \begin{cases} 2^n - 1, & \text{if } S \in \{\emptyset, Q_n\}; \\ 1 + \sum_{x=1}^{|S|} \sum_{y=1}^{n-|S|} \binom{n}{x}\binom{n-x}{y}, & \text{if } \emptyset \subsetneq S \subsetneq Q_n. \end{cases}$$

5. *The star of $L_n(a,b)$ has complexity $2^{n-1} + 2^{n-2}$.*
6. *a) Restricted Complexity:*
   The product $L'_m(a,b)L_n(a,-,b)$ has complexity $m2^n - 2^{n-1}$.

b) Unrestricted Complexity:
The product $L'_m(a,b)L_n(a,c,b)$ has complexity $m2^n + 2^{n-1}$.

7. *a) Restricted Complexity:*
The complexity of $L'_m(a,b) \circ L_n(b,a)$ is mn for $\circ \in \{\cup, \oplus, \setminus, \cap\}$.

   b) Unrestricted Complexity:
The complexity of union and symmetric difference is $mn + m + n + 1$ and this bound is met by $L'_m(a,b,c)$ and $L_n(b,a,d)$, that of difference is $mn+m$ and this bound is met by $L'_m(a,b,c)$ and $L_n(b,a)$, and that of intersection is mn and this bound is met by $L'_m(a,b)$ and $L_n(b,a)$. A total of four letters is required to meet the bounds for union and symmetric difference.

Proof. Clearly $L_n(a)$ has complexity n as the DFA of Definition 1 is minimal.

1. **Syntactic Semigroup** The transformations $a\colon (0,\dots,n-1)$, $b\colon (0,1)$, and $c\colon (n-1 \to 0)$ were used in [5]. It is well known that these transformations as well as a, b, and $c\colon (1 \to 0)$ generate the semigroup of all transformations of Q_n.

2. **Quotients** Obvious.

3. **Reversal** This follows from a theorem in [27] which states that if the transition semigroup has n^n elements, then the complexity of reversal is 2^n. Also, it was shown in [17] that the number of atoms is the same as the complexity of the reverse.

4. **Atoms** Proved in [7, Theorem 3].

5. **Star** Proved in [5].

6. **Product** Let $\mathcal{D}' = (Q'_m, \Sigma', \delta', 0', F')$ and $\mathcal{D} = (Q_n, \Sigma, \delta, 0, F)$ be minimal DFAs of languages L' and L, respectively. We use the standard construction of the ε-NFA $\mathcal{N}$ for the product $L'L$: the final states of $\mathcal{D}'$ becomes non-final, and an ε-transition is added from each state of F' to the initial state 0 of $\mathcal{D}$.

   The subset construction on this NFA yields sets $\{p'\} \cup S$ where $p' \in Q'_m \setminus F'$ and $S \subseteq Q_n$ and sets $\{p', 0\} \cup S$ where $p' \in F'$ and $S \subseteq Q_n \setminus \{0\}$, as well as sets $S \subseteq Q_n$ which can only be reached by letters in $\Sigma \setminus \Sigma'$. Hence the restricted complexity of $L'L$ is bounded by $(m - |F'|)2^n + |F'|2^{n-1} \le m2^n - 2^{n-1}$, and the unrestricted complexity of $L'L$ is bounded by $(m - |F'|)2^n + |F'|2^{n-1} + 2^n \le m2^n + 2^{n-1}$.

   Restricted Complexity: Consider $L'_m(a,b)$ and $L_n(a,-,b)$ of Definition 1; we show that their product meets the upper bound for restricted complexity. As before, we construct an NFA recognizing $L'_m(a,b)L_n(a,-,b)$ and then apply the subset construction to obtain a DFA. Figure 2 shows the NFA for the

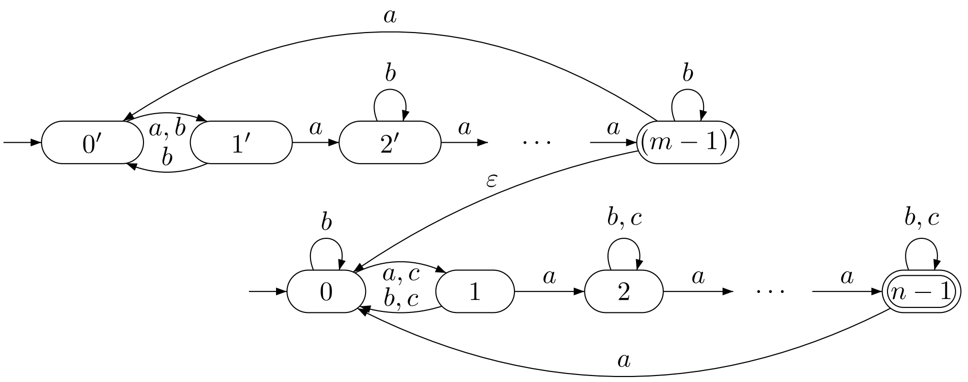

Figure 2: An NFA for the product of $L'_m(a, b)$ and $L_n(a, c, b)$. The NFA for the product of $L'_m(a, b)$ and $L_n(a, -, b)$ is the same except c is omitted.

unrestricted product $L'_m(a, b)L_n(a, c, b)$; the product $L'_m(a, b)L_n(a, -, b)$ is the same except c is omitted.

The initial state is $\{0'\}$ and each state $\{p'\}$ for $0 \le p \le m-2$ is reached by a^p. Consider $\{0'\} \cup S$, where $S = \{q_1, q_2, \dots, q_k\}$ with $0 \le q_1 < q_2 < \cdots < q_k \le n-1$. If $q_1 \ge 1$ then $\{(m-2)', q_2 - q_1 - 1, \dots, q_k - q_1 - 1\} \xrightarrow{a^2} \{0', 1, q_2 - q_1 + 1, \dots, q_k - q_1 + 1\} \xrightarrow{(ab)^{q_1 - 1}} \{0'\} \cup S$. If $q_1 = 0$ and $k \ge 2$, then $\{(m-2)', n-2, q_3 - q_2 - 1, \dots, q_k - q_2 - 1\} \xrightarrow{a^2} \{0', 0, 1, q_3 - q_2 + 1, \dots, q_k - q_2 + 1\} \xrightarrow{(ab)^{q_2 - 1}} \{0'\} \cup S$. State $\{0', 0\}$ is reached by $a^m b^2$. Hence for any non-empty $S \subseteq Q_n$, state $\{0'\} \cup S$ is reachable from $\{(m-2)'\} \cup T$ for some $T \subseteq Q_n$ of size $|S| - 1$. We reach $\{p'\} \cup S$ from $\{0'\} \cup (S - p)$ by a^p, where $S - p$ denotes $\{q - p \mid q \in S\}$ taken mod n. By induction, $\{p'\} \cup S$ is always reachable and thus all $m2^n - 2^{n-1}$ states are reachable.

We check that all states are pairwise distinguishable.

a) Any two sets which differ by $q \in Q_n$ are distinguished by a^{n-1-q}.
b) States $\{p'_1\}$ and $\{p'_2\}$ with $p_1 < p_2$ are distinguished by $a^{m-1-p_2}a^{n-1}$.
c) States $\{0', 0\}$ and $\{p', 0\}$ are distinguished by $(ab)^{m-2-p}aa^{n-1}$ if $p' \ne (m-1)'$; otherwise apply ab to simplify to this case.
d) States $\{p'_1, 0\}$ and $\{p'_2, 0\}$, $p_1 < p_2$, reduce to Case (c) by $(ab^2)^{m-p_2}$.
e) States $\{p'_1\} \cup S$ and $\{p'_2\} \cup S$, where $S \ne \emptyset$ and $p_1 < p_2$, reduce to Case (d) by $(ab)^n$ since $S \xrightarrow{(ab)^n} \{0\}$ and $(ab)^n$ permutes Q'_m.

We can distinguish any pair of states; so the complexity of $L'_m(a, b)L_n(a, -, b)$ is $m2^n - 2^{n-1}$ for all $m, n \ge 3$.

Unrestricted Complexity: The NFA for the product of $L'_m(a, b)L_n(a, c, b)$ is illustrated in Figure 2. The NFA is the same as the restricted case except it has the additional transformation $c\colon (0,1)(Q'_m \to \emptyset)$. Hence the subset construction yields the $m2^n - 2^{n-1}$ sets of the restricted case, as well as all sets $S \subseteq Q_n$ since S is reachable from $\{0'\} \cup S$ by c^2. We check that these sets are distinguishable from all previously reached sets.

a) Any two sets which differ by $q \in Q_n$ are distinguished by a^{n-1-q}.

b) State $\{p'\}$ is distinguishable from $\emptyset$ by $a^{m-1-p}a^{n-1}$.

c) States $\{0',0\}$ and $\{0\}$ are distinguished by $a^{m-1}a^{n-1}$ if $m-1$ is not a multiple of n, and by $ba^{m-2}a^{n-1}$ otherwise.

d) States $\{p',0\}$ and $\{0\}$ reduce to Case (c) by $(ab^2)^{m-p}$.

e) States $\{p'\} \cup S$ and S, where $S \neq \emptyset$, reduce to Case (d) by $(ab)^n$ since $S \xrightarrow{(ab)^n} \{0\}$ and $(ab)^n$ permutes Q'_m.

Hence $L'_m(a,b)L_n(a,c,b)$ has complexity $m2^n + 2^{n-1}$.

7. Boolean Operations

Restricted Complexity: All operations have complexity at most mn [4]. Applying the standard construction for boolean operations we consider the direct product of $\mathcal{D}'_m(a,b)$ and $\mathcal{D}_n(b,a)$ which has states $Q'_m \times Q_n$; the final states vary depending on the operation. By [2, Theorem 1] and computation for the cases $(m,n) \in \{(3,4),(4,3),(4,4)\}$, the states of $Q'_m \times Q_n$ are reachable and pairwise distinguishable for each operation $\circ \in \{\cup, \oplus, \setminus, \cap\}$; hence each operation has complexity mn.

Note that two letters are required to meet these bounds: To a contradiction suppose a single letter ℓ is sufficient to reach $Q'_m \times Q_n$ in the direct product, where $m,n \geq 2$ are not coprime. Letter ℓ must induce an m-element permutation on Q'_m; otherwise there is an unreachable state in Q'_m or the sequence $0', 0'\ell, 0'\ell^2, \ldots, 0'\ell^k, \ldots$ never returns to $0'$. Similarly ℓ must induce an n-cycle in Q_n. Hence ℓ has order $\mathrm{lcm}(mn)$ in the direct product; however, it must have order mn if the bound is to be reached, and this occurs only when m and n are coprime.

Unrestricted Complexity: The upper bounds on the unrestricted complexity of boolean operations are derived in [6]. To compute $L'_m(a,b,c) \circ L_n(b,a,d)$, where $\circ$ is a boolean operation, add an empty state $\emptyset'$ to $\mathcal{D}'_m(a,b,c)$, and send all the transitions from any state of Q'_m under d to $\emptyset'$. Similarly, add an empty state $\emptyset$ to $\mathcal{D}_n(b,a,d)$ together with appropriate transitions; now the alphabets of the resulting DFAs are the same. We consider the direct product of $\mathcal{D}'_{m,\emptyset'}$ and $\mathcal{D}_{n,\emptyset}$ which has states $\{(p',q) \mid p' \in Q'_m \cup \{\emptyset'\}, q \in Q_n \cup \{\emptyset\}\}$. A DFA recognizing $L'_m(a,b,c) \cup L_n(b,a,d)$ is shown in Figure 3 for $m=3$ and $n=4$.

As in the restricted case all the states of $Q'_m \times Q_n$ are reachable by words in $\{a,b\}^*$. The remaining states in $C = \{(p',\emptyset) \mid p' \in Q'_m \cup \{\emptyset'\}\}$ and

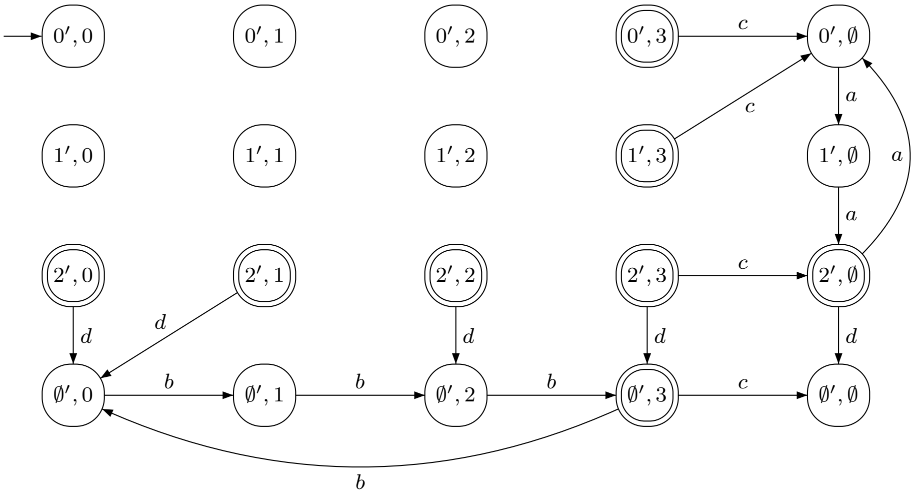

Figure 3: Direct product for union of $\mathcal{D}'_3(a, b, c)$ and $\mathcal{D}_4(b, a, d)$ shown partially.

$R = \{(\emptyset', q) \mid q \in Q_n \cup \{\emptyset\}\}$ are reachable using c and d in addition to a and b as shown in Figure 3. Hence all $(m + 1)(n + 1)$ states are reachable.

For union and symmetric difference, the states of C are pairwise distinguishable by words in a^* and they are distinguished from all other states by words in b^*d. Similarly the states of R are distinguishable from each other and all other states; hence all $mn + m + n + 1$ states are distinguishable.

For difference, the final states are $((m - 1)', q)$ for $q \neq n - 1$. The states of R are all empty, and they are only reachable by d. As the words of $L'_m(a, b, c) \setminus L_n(b, a, d)$ do not contain d, the alphabet is $\{a, b, c\}$; hence we can omit d and delete the states of R, and be left with a DFA recognizing the same language. We check that the remaining $mn+m$ states are pairwise distinguishable. Any states $(p'_1, \emptyset)$ and (p'_2, q) where $p'_1 \neq p'_2$ and $q \in Q_n \cup \{\emptyset\}$ are distinguished by words in a^*. State $(p', \emptyset)$ is distinguished from (p', q) by some $w \in \{a, b\}^*$ that maps (p', q) to $((m - 1)', n - 1)$, since w must send $(p', \emptyset)$ to the final state $((m - 1)', \emptyset)$; such a word exists because a and b induce permutations on the direct product, and so every state in $Q'_m \times Q_n$ is reachable from every other.

For intersection the only final state is $((m - 1)', n - 1)$. The alphabet of $L'_m(a, b, c) \cap L_n(b, a, d)$ is $\{a, b\}$; hence we can omit c and d and delete the states of $R \cup C$, and be left with a DFA recognizing the same language. The remaining mn states are pairwise distinguishable as in the restricted case.

Note that a total of four letters between the alphabets Σ' of $\mathcal{D}'_m$ and Σ of $\mathcal{D}_n$ is required for union and symmetric difference. As in the restricted case, two letters in $\Sigma' \cap \Sigma$ are required to reach the states of $Q'_m \times Q_n$ for general values of m and n. Letters in both alphabets cannot be used to reach states $(p', \emptyset)$ and $(\emptyset', q)$ as the empty states in each coordinate are only reached by letters outside the corresponding alphabet. Thus two additional letters are required, one in $\Sigma' \setminus \Sigma$ and one in $\Sigma \setminus \Sigma'$. Hence each alphabet must contain at least three letters, and $\Sigma' \cup \Sigma$ must contain at least four. In contrast, the bound for difference is met by $L'_m(a,b,c)$ and $L_n(b,a)$, and the bound for intersection is met by $L'_m(a,b)$ and $L_n(b,a)$.

Since all the claims have been verified, the theorem holds. □

5 Right Ideals

The results in this section are based on [8, 9, 18]; however, the stream below is different from that of [18], where c: $(n-2 \to 0)$ and d: $(n-2 \to n-1)$.

Definition 2. *For $n \geq 4$, let $\mathcal{D}_n = \mathcal{D}_n(a,b,c,d) = (Q_n, \Sigma, \delta_n, 0, \{n-1\})$, where $\Sigma = \{a,b,c,d\}$ and δ_n is defined by the transformations a: $(0,\dots,n-2)$, b: $(0,1)$, c: $(1 \to 0)$, and d: $({}_0^{n-2} q \to q+1)$. Let $L_n = L_n(a,b,c,d)$ be the language accepted by $\mathcal{D}_n$. For the structure of $\mathcal{D}_n(a,b,c,d)$ see Figure 4.*

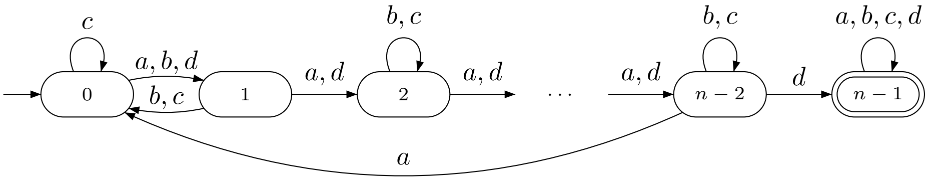

Figure 4: Minimal DFA of a most complex right ideal.

Theorem 2 (Most Complex Right Ideals)**.** *For each $n \geq 4$, the DFA of Definition 2 is minimal and $L_n(a,b,c,d)$ is a right ideal of complexity n. The stream $(L_n(a,b,c,d) \mid n \geq 4)$ with some dialect streams is most complex in the class of right ideals. In particular, it meets all the bounds below. At least four letters are required to meet these bounds.*

1. *The syntactic semigroup of $L_n(a,b,c,d)$ has cardinality n^{n-1}. There is only one maximal transition semigroup of a minimal DFA accepting a right ideal, since it consists of all the transformations of Q_n that fix $n-1$. At least four letters are needed for this bound.*

2. *The quotients of $L_n(a,-,-,d)$ have complexity n, except that $\kappa(\{a,d\}^*) = 1$.*

3. *The reverse of $L_n(a,-,-,d)$ has complexity 2^{n-1}, and $L_n(a,-,-,d)$ has 2^{n-1} atoms.*

4. *Each atom A_S of $L_n(a,b,c,d)$ has maximal complexity:*

$$\kappa(A_S) = \begin{cases} 2^{n-1}, & \text{if } S = Q_n; \\ 1 + \sum_{x=1}^{|S|} \sum_{y=1}^{n-|S|} \binom{n-1}{x-1}\binom{n-x}{y}, & \text{if } \emptyset \subsetneq S \subsetneq Q_n. \end{cases}$$

5. *The star of $L_n(a,-,-,d)$ has complexity $n+1$.*

6. *a) Restricted Complexity:*
   The product $L'_m(a,-,c,d)L_n(a,-,c,d)$ has complexity $m + 2^{n-2}$.

   b) Unrestricted Complexity:
   The product $L'_m(a,-,c,d)L_n(b,-,c,d)$ has complexity $m+2^{n-1}+2^{n-2}+1$. At least three letters for each language and four letters in total are required to meet this bound.

7. *a) Restricted Complexity:*
   The complexity of $\circ$ is mn if $\circ \in \{\cap, \oplus\}$, $mn-(m-1)$ if $\circ = \setminus$, and $mn-(m+n-2)$ if $\circ = \cup$, and these bounds are met by $L'_m(a,-,-,d) \circ L_n(-,-,d,a)$. At least two letters are required to meet these bounds.

   b) Unrestricted Complexity:
   The complexity of $L'_m(a,-,c,d) \circ L_n(b,-,d,a)$ is the same as for arbitrary regular languages: $mn+m+n+1$ if $\circ \in \{\cup, \oplus\}$, $mn+m$ if $\circ = \setminus$, and mn if $\circ = \cap$. At least three letters in each language and four letters in total are required to meet the bounds for intersection and symmetric difference. The bound for difference is also met by $L'_m(a,-,c,d) \setminus L_n(-,-,d,a)$ and the bound for intersection is met by $L'_m(a,-,-,d) \cap L_n(-,-,d,a)$.

Proof. DFA $\mathcal{D}_n(-,-,-,d)$ is minimal because the shortest word in d^* accepted by state q is d^{n-1-q}, and $L_n(a,b,c,d)$ is a right ideal because it has only one final state and that state accepts Σ^*.

1. **Semigroup** The transformations induced by a, b, and c generate all transformations of Q_{n-1}. Also, since the transformation induced by da^{n-2} is $(n-2 \to n-1)$, the transition semigroup of $\mathcal{D}_n(a,b,c,d)$ contains the one in [18], which is maximal for right ideals. Hence the syntactic semigroup of $L_n(a,b,c,d)$ has size n^{n-1} as well. The fact that at least four letters are needed was proved in [15].

2. **Quotients** If the initial state of $\mathcal{D}_n(a,-,-,d)$ is changed to q with $0 \le q \le n-2$, the new DFA accepts a quotient of L_n and is still minimal; hence the complexity of that quotient is n.

3. **Reversal** It was proved in [10] that the reverse has complexity 2^{n-1}, and in [17] that the number of atoms is the same as the complexity of the reverse.

4. **Atoms** The proof in [8] applies since the DFA has all the transformations that fix $n-1$.

5. **Star** If L_n is a right ideal, then $L_n^* = L_n \cup \{\varepsilon\}$. If we add a new initial state $0'$ to the DFA of Definition 2 with the same transitions as those from 0 and make $0'$ final, the new DFA accepts L_n^* and is minimal for $0'$ does not accept a, and so is not equivalent to $n-1$.

6. **Product** Let $\mathcal{D}' = (Q'_m, \Sigma', \delta', 0', \{(m-1)'\})$ and $\mathcal{D} = (Q_n, \Sigma, \delta, 0, \{n-1\})$ be minimal DFAs of L' and L, respectively, where L' and L are right ideals. We use the standard construction of the NFA for the product $L'L$: the final state $(m-1)'$ of $\mathcal{D}'$ becomes non-final, and an ε-transition is added from that state to the initial state 0 of $\mathcal{D}$. We bound the complexity of the product by counting the reachable states in the subset construction on this NFA. The $m-1$ non-final states $\{p'\}$ of $\mathcal{D}'$ may be reachable, as well as $\{(m-1)', 0\}$. From $\{(m-1)', 0\}$ we may reach all 2^{n-2} subsets of Q_n which contain 0 but not $n-1$, and 2^{n-2} states that contain both 0 and $n-1$; however, the latter 2^{n-2} states all accept Σ^* and are therefore equivalent. So far, we have $m - 1 + 2^{n-2} + 1 = m + 2^{n-2}$ states; these are the only reachable sets if the witnesses are restricted to the same alphabet.

   For the unrestricted case, suppose that $\ell' \in \Sigma' \setminus \Sigma$ and $\ell \in \Sigma \setminus \Sigma'$. By applying ℓ to $\{(m-1)', 0\} \cup S$, $S \subseteq Q_n \setminus \{0\}$, we may reach all $2^n - 1$ nonempty subsets of Q_n, and then by applying ℓ' we reach the empty subset. However, the 2^{n-1} subsets of Q_n that contain $n-1$ all accept Σ^*. Hence there are at most $2^{n-1}+1$ additional sets, for a total of $m + 2^{n-2} + 2^{n-1} + 1$ reachable sets.

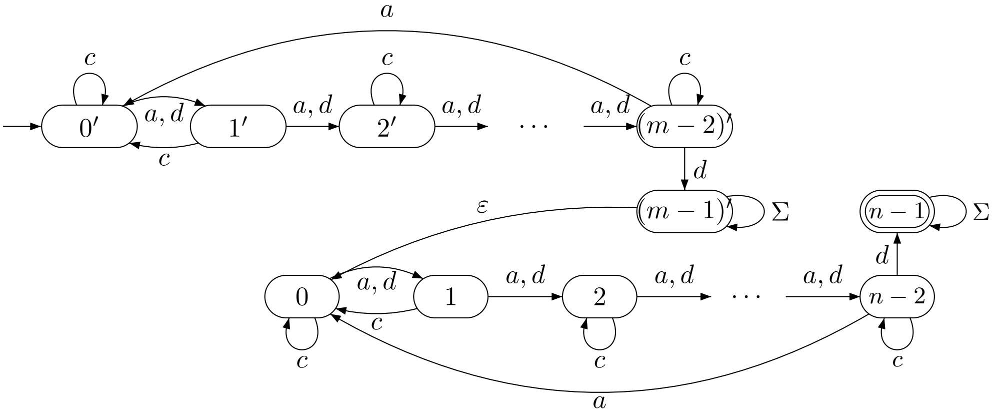

Figure 5: An NFA for product of right ideals $L'_m(a, -, c, d)$ and $L_n(a, -, c, d)$.

Restricted Complexity: To prove the bound is tight, consider the two dialects of the DFA of Definition 2 shown in Figure 5, where $\Sigma = \{a, c, d\}$. The

$m-1$ sets $\{p'\}$ for $p' \in Q'_{m-1}$ are reachable in $\mathcal{D}'_m$ by words in d^*, and $\{(m-1)', 0\}$ is reached by d^{m-1}. The 2^{n-2} sets of the form $\{(m-1)', 0\} \cup S$, where $S \subseteq Q_n \setminus \{0\}$, are reachable using words in $\{c, d\}^*$ as follows: To reach $\{(m-1)', 0\} \cup S$, where $S = \{q_1, \dots, q_k\}$, $1 \le q_1 < q_2 < \cdots < q_k \le n-1$, we have first $\{(m-1)', 0\}d = \{(m-1)', 0, 1\}$. State 1 will then be moved to the right by applying either d or dc repeatedly: If $q_{k-1} = q_k - 1$, use d; otherwise use dc. Repeating this process q_k times we eventually construct S. For example, to reach $\{(m-1)', 0\} \cup \{2, 5, 7, 8\}$ use $dd(dc)d(dc)(dc)d(dc)$. The 2^{n-2} sets $\{(m-1)', 0\} \cup S$ that contain $n-1$ all accept $\{a, c, d\}^*$; hence they are all equivalent.

The remaining states are pairwise distinguishable: States $\{p'\}$ and $\{q'\}$ with $0 \le p < q \le m-2$ are distinguished by $d^{m-1-q}d^{n-1}$, and $\{p'\}$ is distinguished from $\{(m-1)', 0\} \cup S$ by d^{n-1}. Two non-final states $\{(m-1)', 0\} \cup S$ and $\{(m-1)', 0\} \cup T$ with $q \in S \oplus T$ are distinguished by $a^{n-2-q}d$. Thus the product has complexity $m + 2^{n-2}$.

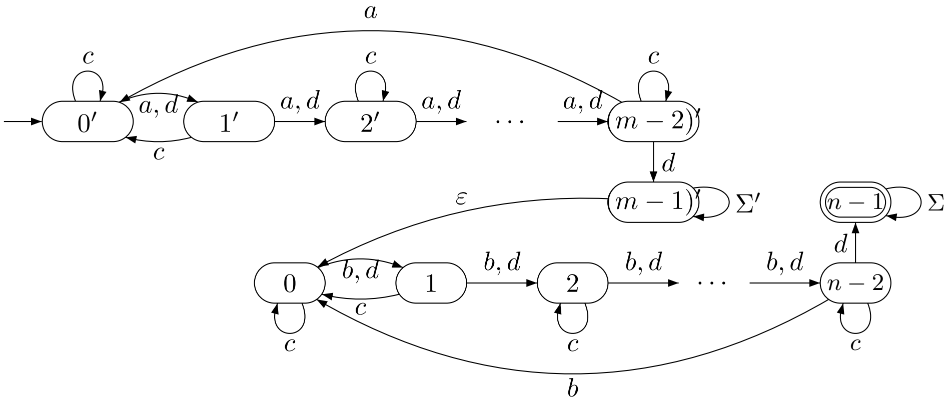

Figure 6: An NFA for product of right ideals $L'_m(a, -, c, d)$ and $L_n(b, -, c, d)$.

Unrestricted Complexity: Consider two dialects of the DFA of Definition 2 shown in Figure 6. Here $\Sigma' = \{a, c, d\}$ and $\Sigma = \{b, c, d\}$. By the restricted case, all states $\{p'\}$ for $p' \in Q'_{m-1}$ and $\{(m-1)', 0\} \cup S$ for $S \subseteq Q_n \setminus \{0\}$ are reachable by words in $\{c, d\}^*$. Apply b from $\{0'\}$ to reach the empty subset. By applying b to $\{(m-1)', 0\} \cup S$, $S \subseteq Q_n \setminus \{0\}$, we reach all $2^n - 1$ non-empty subsets of Q_n; hence all states are reachable. However, the 2^{n-1} sets $S \subseteq Q_n$ that contain $n-1$ all accept $\{b, c, d\}^*$ and are sent to the empty state by a; hence they are all equivalent. Similarly, the 2^{n-2} sets $\{(m-1)', 0\} \cup S$ that contain $n-1$ all accept $\{b, c, d\}^*$ and are sent to $\{(m-1)', 0\}$ by a; hence they are also equivalent.

The remaining states are pairwise distinguishable. States $\{p'\}$ and $\{q'\}$ with $0 \le p < q \le m-2$ are distinguished by $d^{m-1-q}d^{n-1}$, and $\{p'\}$ is distinguished from $\{(m-1)', 0\} \cup S$ or from S, where $\emptyset \subsetneq S \subseteq Q_n$, by d^{n-1}. Two states

$\{(m-1)', 0\} \cup S$ and $\{(m-1)', 0\} \cup T$ with $q \in S \oplus T$ are distinguished by $b^{n-2-q}d$, as are two states S and T with $q \in S \oplus T$. A state $\{(m-1)', 0\} \cup S$ is distinguishable from T where $S, T \subseteq Q_n$ by ad^{n-1}. Thus all $m + 2^{n-2} + 2^{n-1} + 1$ states are pairwise distinguishable.

At least three inputs to each DFA are required to achieve the bound in the unrestricted case: There must be a letter in Σ (like d) with a transition to $n-1$ to reach sets containing $n-1$, and this letter must be in Σ' in order to reach the sets that contain both $(m-1)'$ and $n-1$. However no single letter in $\Sigma' \cap \Sigma$ is sufficient to reach every set of the form $\{(m-1)', 0\} \cup S$, regardless of its behaviour on Q_n. For example, if the letter maps $0 \to q_1$ and $q_1 \to q_2$ then it is impossible to reach the state $\{(m-1)', 0, q_2\}$ by repeatedly applying the letter from $\{(m-1)', 0\}$, as it can never delete q_1. Hence there must be at least two letters in $\Sigma' \cap \Sigma$. Furthermore there must be some $\ell \in \Sigma \setminus \Sigma'$ to reach the empty state, and there must be some $\ell' \in \Sigma' \setminus \Sigma$ to distinguish $\{(m-1)', 0, n-1\}$ from $\{n-1\}$. Thus each alphabet must contain at least three letters to meet the bound.

7. Boolean Operations

Restricted Complexity: The bounds for right ideals were derived in [10]. We show that the DFAs $\mathcal{D}'_m(a, -, -, d)$ and $\mathcal{D}_n(-, -, d, a)$ shown in Figure 7 of the right ideals of Definition 2 meet the bounds.

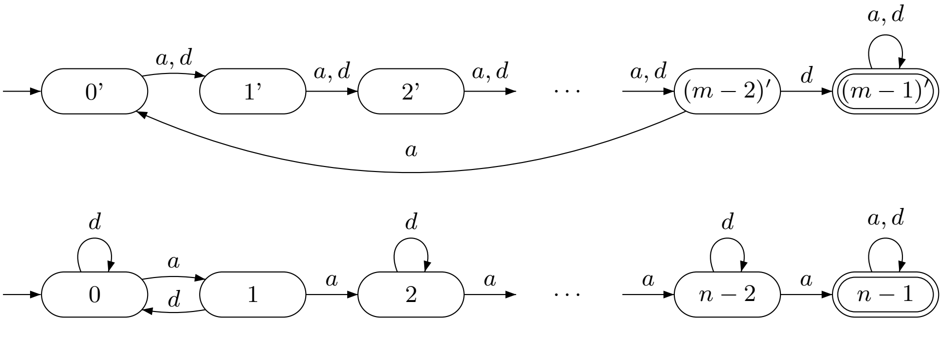

Figure 7: DFAs of $L'_m(a, -, -, d)$ and $L_n(-, -.d, a)$ for boolean operations.

Consider the direct product of $L'_m(a, -, -, d)$ and $L_n(-, -, d, a)$, illustrated in Figure 8 for $m = n = 4$. For $p' \in Q'_{m-1}$ state $(p', 0)$ is reached by d^p. Since the first column of $Q_{m-1} \times Q_n$ is reachable and $(p', q) \xrightarrow{a} ((p+1)', (q+1))$, where $p+1$ is taken mod $m-1$, we can reach every state in $Q_{m-1} \times Q_n$. State $((m-1)', q)$ is reached by $d^{m-1}a^q$; hence the states of $Q'_m \times Q_n$ are reachable.

We now check distinguishability, which depends on the final states of the

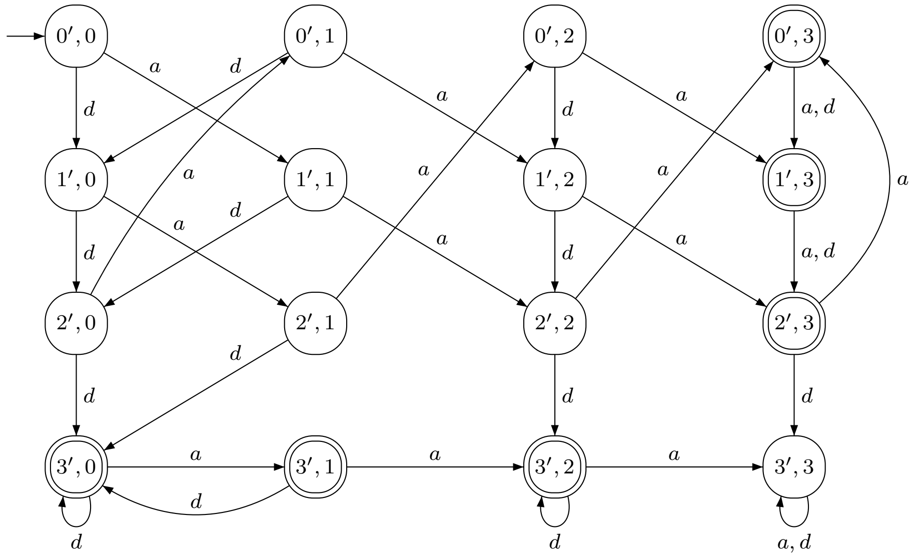

Figure 8: Partial illustration of direct product for $L'_4(a, -, -, d) \oplus L_4(-, -, d, a)$.

DFA. The direct product is made to recognize $L'_m(a, -, -, d) \circ L_n(-, -, d, a)$ by setting the final states to be $(\{(m-1)'\} \times Q_n) \circ (Q'_m \times \{n-1\})$.

For intersection and symmetric difference, all states are pairwise distinguishable. States that differ in the first coordinate are distinguished by words in d^*a^* and states that differ in the second coordinate are distinguished by words in a^*d^*. Hence the complexity is mn.

For difference, the states $\{(p', n-1) \mid p' \in Q'_m\}$ are all empty, and therefore equivalent. The remaining states are non-empty, and they are distinguished by words in d^* if they differ in the first coordinate or by words in a^*d^* if they differ in the second coordinate. Hence the complexity is $mn - m + 1$.

For union, the states $\{(p', n-1) \mid p' \in Q'_m\} \cup \{((m-1)', q) \mid q \in Q_n\}$ are all final and equivalent as they accept $\{a, d\}^*$. The remaining states are distinguished by words in d^* if they differ in the first coordinate or by words in a^* if they differ in the second coordinate. Hence the complexity is $mn - (m + n - 2)$.

As in regular languages, one letter in $\Sigma' \cap \Sigma$ is not sufficient to reach all the states of $Q'_{m-1} \times Q_{n-1}$ for all values of m and n; hence two letters are required to meet any of the bounds.

Unrestricted Complexity: The unrestricted bounds for right ideals are the same as those for arbitrary regular languages [6]. We show that the DFAs $\mathcal{D}'_m(a, -, c, d)$ and $\mathcal{D}_n(b, -, d, a)$ of Definition 2 meet the bounds.

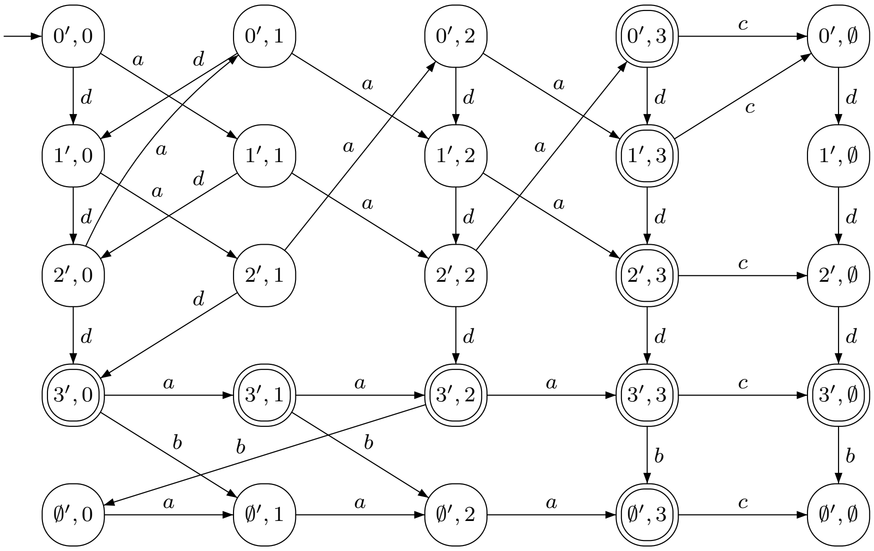

Figure 9: Partial illustration of the direct product for $L'_4(a,-,c,d) \cup L_4(b,-,d,a)$.

To compute $L'_m(a,-,c,d) \circ L_n(b,-,d,a)$, where $\circ$ is a boolean operation, add an empty state $\emptyset'$ to $\mathcal{D}'_m(a,-,c,d)$, and send all the transitions from any state of Q'_m under b to $\emptyset'$. Similarly, add an empty state $\emptyset$ to $\mathcal{D}_n(b,-,d,a)$ together with appropriate transitions; now the alphabets of the resulting DFAs are the same. The direct product of $L'_m(a,-,c,d)$ and $L_n(b,-,d,a)$ is illustrated in Figure 9 for $m = n = 4$.

As in the restricted case, the mn states of $Q'_m \times Q_n$ are reachable by words in $\{a,d\}^*$. The remaining states $(p',\emptyset)$ and $(\emptyset',q)$ are easily seen to be reachable using b and c, as well as a and d.

We now check distinguishability, which depends on the final states of the DFA. The direct product is made to recognize $L'_m(a,-,c,d) \circ L_n(b,-,d,a)$ by setting the final states to be $(\{(m-1)'\} \times Q_n \cup \{\emptyset\}) \circ (Q'_m \cup \{\emptyset'\} \times \{n-1\})$.

For union and symmetric difference, all states are pairwise distinguishable: States that differ in the first coordinate are distinguished by words in d^*c and states that differ in the second coordinate are distinguished by words in a^*b.

For difference, the final states are $((m-1)',q)$ for $q \neq n-1$. The alphabet of $L'_m(a,-,c,d) \setminus L_n(b,-,a,d)$ is $\{a,c,d\}$; hence we can omit b and delete all states $(\emptyset',q)$ and be left with a DFA recognizing the same language. The remaining states are distinguished by words in d^*c if they differ in the first coordinate or by words in a^*d^* if they differ in the second coordinate.

For intersection, the only final state is $((m-1)',n-1)$. The alphabet of $L'_m(a,-,c,d) \cap L_n(b,-,d,a)$ is $\{a,b\}$; hence we can omit b and c and delete all states $(p', \emptyset)$ and $(\emptyset', q)$. The remaining mn states are pairwise distinguishable as in the restricted case.

Note that the bound for difference is met by $L'_m(a, -, c, d) \setminus L_n(-, -, d, a)$, and that of intersection is met by $L'_m(a, -, -, d) \cap L_n(-, -, d, a)$. However the bounds for union and symmetric difference all require three letters in each dialect: There must be a letter in $\Sigma' \setminus \Sigma$ to reach states of the form $(p', \emptyset)$, and there must a letter in $\Sigma \setminus \Sigma'$ to reach states of the form $(\emptyset', q)$. As in regular languages, one letter in $\Sigma' \cap \Sigma$ is not sufficient to reach all the states of $Q'_m \times Q_n$ for all values of m and n; hence $|\Sigma' \cap \Sigma| \geq 2$ and so both Σ' and Σ must contain at least three letters.

It has been shown in [10] that at least two letters are needed for each right ideal that meets the bounds for star or reversal. Hence almost all our witnesses in Theorem 2 that meet the bounds for the common operations use minimal alphabets. □

6 Prefix-Closed Languages

The complexity of operations on prefix-closed languages was studied in [11], but most complex prefix-closed languages were not considered. As every prefix-closed language has an empty quotient, the restricted and unrestricted complexities are the same for binary operations.

Definition 3. *For $n \geq 4$, let $\mathcal{D}_n = \mathcal{D}_n(a, b, c, d) = (Q_n, \Sigma, \delta_n, 0, Q_n \setminus \{n-1\})$, where $\Sigma = \{a, b, c, d\}$, and δ_n is defined by the transformations $a\colon (0, \dots, n-2)$, $b\colon (0, 1)$, $c\colon (1 \to 0)$, and $d\colon \binom{0}{n-2}\, q \to q - 1 \pmod n$. Let $L_n = L_n(a, b, c, d)$ be the language accepted by $\mathcal{D}_n$. The structure of $\mathcal{D}_n(a, b, c, d)$ is shown in Figure 10.*

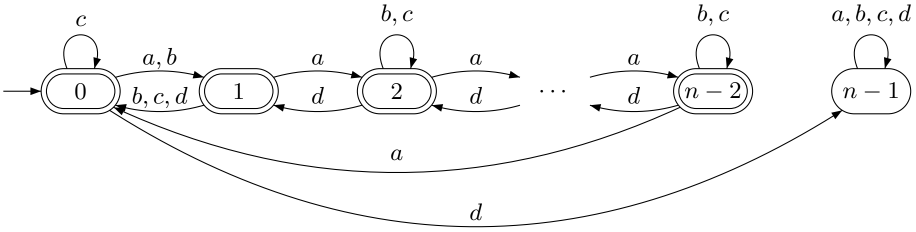

Figure 10: DFA of a most complex prefix-closed language.

Theorem 3 (Most Complex Prefix-Closed Languages)**.** *For each $n \geq 4$, the DFA of Definition 3 is minimal and $L_n(a, b, c, d)$ is a prefix-closed language of complexity n. The stream $(L_m(a, b, c, d) \mid m \geq 4)$ with some dialect streams is most complex in the class of prefix-closed languages. At least four letters are required to meet the bounds below.*

1. *The syntactic semigroup of $L_n(a,b,c,d)$ has cardinality n^{n-1}.*
2. *The quotients of $L_n(a,-,-,d)$ have complexity n, except for $\emptyset$, which has complexity 1.*
3. *The reverse of $L_n(a,-,-,d)$ has complexity 2^{n-1}, and $L_n(a,-,-,d)$ has 2^{n-1} atoms.*
4. *Each atom A_S of $L_n(a,b,c,d)$ has maximal complexity:*

$$\kappa(A_S) = \begin{cases} 2^{n-1}, & \textit{if } S = \emptyset; \\ 1 + \sum_{x=1}^{n-|S|} \sum_{y=1}^{|S|} \binom{n-1}{x-1}\binom{n-x}{y}, & \textit{if } \emptyset \subsetneq S \subsetneq Q_n. \end{cases}$$

5. *The star of $L_n(a,-,c,d)$ has complexity $2^{n-2}+1$.*
6. *The product $L'_m(a,b,c,d)L_n(a,d,b,c)$ has complexity $(m+1)2^{n-2}$.*
7. *For any proper binary boolean function $\circ$, the complexity of $L'_m(a,b,-,d) \circ L_n(b,a,-,d)$ is maximal. In particular, the complexity is mn if $\circ \in \{\cup, \oplus\}$, $mn-(n-1)$ if $\circ = \setminus$, and $mn-(m+n-2)$ if $\circ = \cap$.*

Proof. The DFA is minimal since state p rejects d^q if and only if $p < q$. It is prefix-closed because all non-empty states are final.

1. **Semigroup** Let d' induce the transformation $({}_{0}^{n-2} q \to q+1)$ (this was called d in the right-ideal section). Since $ada = d'$, the transition semigroup of the DFA of Figure 10 is the same as that of the DFA of the right ideal of Figure 4.
2. **Quotients** Obvious.
3. **Reversal** Since reversal commutes with complementation, we consider the complement of the language accepted by the DFA of Figure 10 restricted to the alphabet $\{a,d\}$. It was proved in [10] that the reverse of a right ideal has complexity at most 2^{n-1}, and in [17] that the number of atoms is the same as the complexity of the reverse. It remains to prove that all 2^{n-1} states of the DFA $\mathcal{D}^R$ obtained by the subset construction from the NFA $\mathcal{N}$ obtained by reversal of the DFA of the right ideal $\mathcal{D}$ are reachable and distinguishable. The proof is similar to that of [10]. Subset $\{n-1\}$ is the initial state of $\mathcal{N}$, and $n-1$ appears in every reachable state of $\mathcal{D}^R$. Every subset $\{n-1, q_2, q_3 \dots, q_k\}$ of size k, where $1 \le k \le n-2$ and $0 \le q_2 < q_3 < \cdots < q_k \le n-2$, is reached from the subset $\{n-1, q_3-(q_2+1), \dots, q_k-(q_2+1)\}$ of size $k-1$ by $da^{n-(q_2+1)}$. Since only state q, $0 \le q \le n-2$, accepts a^q, any two subsets differing by q are distinguishable by a^q.
4. **Atoms** Let L be a prefix-closed language with quotients $K_0, \dots, K_{n-1}$, $n \ge 4$. Recall that $\overline{L}$ is a right ideal with quotients $\overline{K_0}, \dots, \overline{K_{n-1}}$. For $S \subseteq \{0, \dots, n-1\}$, the atom of L corresponding to S is $A_S = \bigcap_{i \in S} K_i \cap \bigcap_{i \in \overline{S}} \overline{K_i}$. This can be rewritten as $\bigcap_{i \in \overline{S}} \overline{K_i} \cap \bigcap_{i \in \overline{\overline{S}}} \overline{\overline{K_i}}$, which is the atom of $\overline{L}$ corresponding

to $\overline{S}$; hence the sets of atoms of L and $\overline{L}$ are the same. The claim follows from the theorem for right ideals. The proof in [8] applies since the DFA that accepts the complement of the prefix-closed language of Figure 10 has all the transformations that fix $n-1$.

5. **Star** It was proved in [11] that $2^{n-2}+1$ is the maximal complexity of the star of a prefix-closed language. We now show that $L_n(a,-,c,d)$ meets this bound. Since $L_n(a,-,c,d)$ accepts ε, no new initial state is needed and it suffices to delete the empty state and add an ε-transition from each final state to the initial state to get an NFA $\mathcal{N}$ for L_n^*. In this NFA all 2^{n-2} subsets of Q_{n-1} containing 0 are reachable and pairwise distinguishable. Any non-empty set $\{0,q_2,q_3,\dots,q_k\}$ of size k with $0<q_2<q_3<\dots<q_k\le n-2$ is reached from $\{0,q_3-q_2,\dots,q_k-q_2\}$ of size $k-1$ by $a(ac)^{q_2-1}$. Moreover, the empty set is reached from $\{0\}$ by d, giving the required bound. Sets $\{0\}\cup S$ and $\{0\}\cup T$ with $q\in S\oplus T$ are distinguished by $a^{n-2-q}d^{n-2}$.

6. **Product** It was shown in [11] that the complexity of the product of prefix-closed languages is $(m+1)2^{n-2}$. We now prove that our witness $L'_m(a,b,c,d)$ with minimal DFA $\mathcal{D}'_m(a,b,c,d)$ together with the dialect $L_n(a,d,b,c)$ with minimal DFA $\mathcal{D}_n(a,d,b,c)$ meets this bound. Construct the following NFA $\mathcal{N}$ for the product. Start with $\mathcal{D}'_m(a,b,c,d)$, but make all of its states non-final. Delete the empty state from $\mathcal{D}_n(a,d,b,c)$ and all the transitions to the empty state, add an ε-transition from each state $p'\in Q'_{m-1}$ to the initial state 0 of $\mathcal{D}_n(a,d,b,c)$. We will show that $(m-1)2^{n-2}$ states of the form $\{p',0\}\cup S$, where $S\subseteq Q_{n-1}\setminus\{0\}$, and 2^{n-1} states of the form $\{(m-1)'\}\cup S$, where $S\subseteq Q_{n-1}$ are reachable and pairwise distinguishable.

   The initial state of the subset automaton of $\mathcal{N}$ is $\{0',0\}$. State $\{1',0\}$ is reachable by b and $\{p',0\}$ for $2\le p\le m-2$ is reachable from $\{1',0\}$ by $(ab)^{p-1}$. State $\{p',0\}\cup S$ where $p'\in Q'_{m-1}$ and $S=\{q_1,\dots,q_k\}$ is reachable from $\{r',0,q_2-q_1,\dots,q_k-q_1\}$ by $a(ab)^{q_1-1}$ for some $r'\in Q'_{m-1}$. By induction, all $(m-1)2^{n-2}$ states $\{p',0\}\cup S$ are reachable. From $\{0',0\}\cup S$ by d^2 we reach $\{(m-1)',0\}\cup S$. Further apply ca to reach $\{(m-1)'\}\cup S$. Hence all 2^{n-1} subsets of the form $\{(m-1)'\}\cup S$ are reachable.

   We check that the states are pairwise distinguishable in four cases.

   a) $\{(m-1)'\}\cup S$ and $\{(m-1)'\}\cup T$ with $r\in S\oplus T$ are distinguished by $a^{n-2-r}c^{n-2}$.

   b) $\{p'\}\cup S$ and $\{p'\}\cup T$ with $r\in S\oplus T$ reduces to Case (a) by $a^{n-2-r}d^m$.

   c) $\{p'\}\cup S$ and $\{(m-1)'\}\cup T$ with $p'\in Q'_{m-1}$ are distinguished by c^n.

   d) $\{p'\}\cup S$ and $\{q'\}\cup T$ with $p<q\le m-2$ reduces to Case (c) by d^{p+1}.

7. **Boolean Operations** It is again convenient to consider the ideal languages defined by the complements of the prefix-closed languages of Figure 10 restricted to the alphabet $\{a,b,d\}$ and then use De Morgan's laws. Since every prefix-closed language has an empty quotient, it is sufficient to consider

boolean operations on languages over the same alphabet. The problems are the same as those in [9], except that there the transformation induced by d is $d : (n-2 \to n-1)$.

Let $\mathcal{D}_n(a,b,c,d)$ denote the DFA for the complement of the prefix-closed language of Definition 3 of complexity n and let L_n be the language accepted by $\mathcal{D}_n$. We consider boolean operations on right ideals L'_m and L_n.

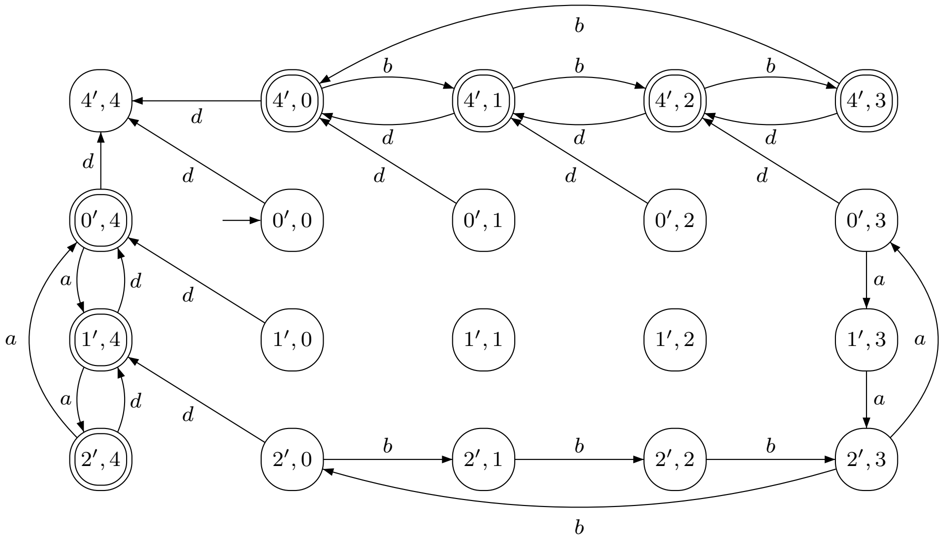

Figure 11: Partial illustration of the direct product for $L'_4(a,b,-,d) \oplus L_5(b,a,-,d)$.

The direct product is illustrated in Figure 11. The states in $Q'_{m-1} \times Q_{n-1}$ are reachable from the initial state $(0',0)$ by [2, Theorem 1] and computation in the case $m = n = 4$. Then $((m-1)',0)$ is reached from $(0',1)$ by d and states of the form $((m-1)',q)$, $0 \le q \le n-2$, are then reached by words in b^*. Similarly, $(0',n-1)$ is reached from $(1',0)$ by d and states of the form $(p',n-1)$, $0 \le p \le m-2$, are then reached by words in a^*. Finally, $((m-1)',n-1)$ is reached from $((m-1)',0)$ by d. Hence all mn states are reachable.

Let $S = Q'_{m-1} \times Q_{n-1}$, $R = \{(m-1)'\} \times Q_n$, and $C = Q'_m \times \{n-1\}$. The final states of the direct product to recognize $L'_m(a,b,-,d) \circ L_n(b,a,-,d)$ are $R \circ C$.

Consider the following DFAs: $D'_{m-1}(a,b) = (Q'_{m-1}, \{a,b\}, \delta, 0', \{0'\})$ and $D_{n-1}(b,a) = (Q_{n-1}, \{a,b\}, \delta, 0, \{0\})$. By [2, Theorem 1], the states of S are pairwise distinguishable with respect to states $(\{0'\} \times Q_{n-1}) \circ \big(Q'_{m-1} \times \{0\}\big)$ for any $\circ \in \{\cup, \oplus, \setminus, \cap\}$. One can verify that if w distinguished two states of S with respect to $(\{0'\} \times Q_{n-1}) \circ \big(Q'_{m-1} \times \{0\}\big)$, then wd distinguishes them with respect to $R \circ C$ for each $\circ = \{\cup, \oplus, \setminus, \cap\}$. The rest of the argument depends on the operation $\circ \in \{\cup, \oplus, \setminus, \cap\}$.

$\cap, \oplus$: All mn states are pairwise distinguishable. The states of R are distinguished by words in d^*. The states of C are similarly distinguishable. The states of R are distinguished from the states of C by words in $\{a,d\}^*$. Every state of S is sent to a state of R by a word in $\{a,d\}^*$, and similarly to a state of C by a word in $\{b,d\}^*$; thus the states of S are distinguishable from the states of R or C.

$\setminus$: The states of C are all empty, leaving $m(n-1)+1$ distinguishable states. The states of R are distinguished by words in d^*.

$\cup$: The states of R and C are equivalent final states accepting all words, leaving $(m-1)(n-1)+1$ distinguishable states.

By De Morgan's laws we have $\kappa(L'_m \cup L_n) = \kappa(\overline{L'_m} \cap \overline{L_n})$, $\kappa(L'_m \oplus L_n) = \kappa(\overline{L'_m} \oplus \overline{L_n})$, $\kappa(L'_m \setminus L_n) = \kappa(\overline{L_n} \setminus \overline{L'_m})$, and $\kappa(L'_m \cap L_n) = \kappa(\overline{L'_m} \cup \overline{L_n})$. Thus the prefix-closed witness meets the bounds for boolean operations.

Since the semigroup of a prefix-closed language is the same as that of its complement, which is a right ideal, at least four letters are required to meet all the bounds. □

7 Prefix-Free Languages

The complexity of operations on prefix-free languages was studied in [19, 22, 23], but most complex prefix-free languages were not considered. As every prefix-free language has an empty quotient, the restricted and unrestricted complexities are the same for binary operations.

Definition 4. *For $n \ge 4$, let $\Sigma_n = \{a,b,c,d,e_0,\dots,e_{n-3}\}$ and define the DFA $\mathcal{D}_n(\Sigma_n) = (Q_n, \Sigma_n, \delta_n, 0, \{n-2\})$, where δ_n is defined by the transformations $a\colon (n-2 \to n-1)(0,\dots,n-3)$, $b\colon (n-2 \to n-1)(0,1)$, $c\colon (n-2 \to n-1)(1 \to 0)$, $d\colon (0 \to n-2)(Q_n \setminus \{0\} \to n-1)$, $e_q\colon (n-2 \to n-1)(q \to n-2)$ for $q = 0,\dots,n-3$. The transformations induced by a and b coincide when $n=4$. Let $L_n(\Sigma_n)$ be the language accepted by $\mathcal{D}_n(\Sigma_n)$. The structure of $\mathcal{D}_n(\Sigma_n)$ is shown in Figure 12.*

Theorem 4. *For $n \ge 4$, the DFA of Definition 4 is minimal and $L_n(\Sigma)$ is a prefix-free language of complexity n. The stream $(L_n(a,b,c,d,e_0,\dots,e_{n-3}) \mid n \ge 4)$ with dialect stream $(L_n(b,a,-,-,e_0,\ e_{n-3}) \mid n \ge 4)$ is most complex in the class of prefix-free languages. At least $n+2$ inputs are required to meet all the bounds below.*

1. *The syntactic semigroup of $L_n(a,b,c,-,e_0,\dots,e_{n-3})$ has cardinality n^{n-2}. There is only one maximal transition semigroup of minimal DFAs accepting prefix-free languages. Moreover, fewer than $n+1$ inputs do not suffice to meet this bound.*

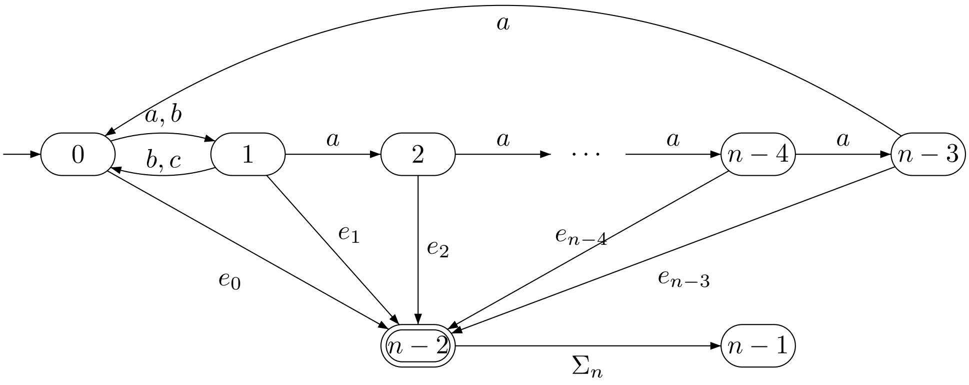

Figure 12: DFA of a most complex prefix-free language. Input d not shown; other missing transitions are self-loops.

2. *The quotients of $L_n(a,-,-,d)$ have complexity n, except for ε and $\emptyset$, which have complexity 2 and 1, respectively.*

3. *The reverse of $L_n(a,-,c,-,e_0)$ has complexity $2^{n-2}+1$, and $L_n(a,-,c,-,e_0)$ has $2^{n-2}+1$ atoms.*

4. *Each atom A_S of $L_n(a,b,c,-,e_0)$ has maximal complexity:*

$$\kappa(A_S) = \begin{cases} 2, & \text{if } S = \{n-2\}; \\ 2^{n-1}, & \text{if } S = \emptyset; \\ 2^{n-2}+1, & \text{if } S = Q_{n-2}; \\ 2 + \sum_{x=1}^{|S|} \sum_{y=1}^{n-2-|S|} \binom{n-2}{x}\binom{n-2-x}{y}, & \text{if } \emptyset \subsetneq S \subsetneq Q_{n-2}. \end{cases}$$

5. *The star of $L_n(a,-,-,d)$ has complexity n.*

6. *The product $L'_m(a,-,-,d)L_n(a,-,-,d)$ has complexity $m+n-2$.*

7. *For $m,n \ge 4$ but $(m,n) \ne (4,4)$, and for any proper binary boolean function $\circ$, the complexity of $L_m(a,b,-,-,e_0,e_{m-3}) \circ L_n(b,a,-,-,e_0,e_{n-3})$ is maximal. In particular, these languages meet the bounds $mn-2$ for union and symmetric difference, $mn-2(m+n-3)$ for intersection, and $mn-(m+2n-4)$ for difference.*

Proof. Since only state q accepts $a^{n-2-q}d$ for $0 \le q \le n-3$, DFA $\mathcal{D}_n(a,-,-,d)$ is minimal. Since it has only one final state and that state accepts $\{\varepsilon\}$, $L_n(a,-,-,d)$ is prefix-free.

1. **Semigroup** The proof that the size of the semigroup is n^{n-2} is very similar to that in [12]. Inputs a, b, and c generate all transformations of Q_{n-2}. Moreover, any state $q \in Q_{n-2}$ can be sent to $n-2$ by e_q and to $n-1$ by e_qe_q.

Hence we have all n^{n-2} transformations of Q_n that fix $n-1$ and send $n-2$ to $n-1$. The maximal transition semigroup is unique, since it must contain all these transformations.

To prove that at least $n+1$ inputs are necessary, we see that $e_q\colon (n-2 \to n-1)(q \to n-2)$ is in the transition semigroup of $\mathcal{D}_n$. There are two types of states in $q \in Q_{n-2}$: those of Type 1, for which e_q is a generator (that is, the transformation of e_q is induced by a single letter), and those of Type 2, for which it is not. If e_q and e_p are generators, then clearly $e_p \neq e_q$.

If e_q is not a generator, then it must be a composition, $e_q = u_q v_q$, where u_q is in the semigroup and v_q is a generator. No state can be mapped by u_q to $n-2$ because then v_q would map $n-2$ to $n-1$. Hence u_q must be a permutation of Q_{n-2}. If $q \neq q'$ and e_q and $e_{q'}$ are not generators, then there exist u_q, v_q and $u_{q'}, v_{q'}$ as above, such that $e_q = u_q v_q$ and $e_{q'} = u_{q'} v_{q'}$. Then we must have $qu_q \neq q'u_{q'}$; otherwise both q and q' would be mapped to $n-2$. Hence $v_q \neq v_{q'}$ and all the generators of this type are distinct.

Finally, if e_q is a generator and $v_{q'}$ is as above, then $e_q \neq v_{q'}$, for otherwise $u_{q'}$ would be the identity and q' would be of Type 1. Therefore, $n-2$ generators are required in addition to those induced by a, b and c.

2. **Quotients** This is clear from the definition.

3. **Reversal** This was proved in [12].

4. **Atoms** First we establish an upper bound on the complexity of atoms of prefix-free languages. Let L be a prefix-convex language with n quotients $K_0, \dots, K_{n-1}$, in which K_{n-2} is final and $K_{n-1} = \emptyset$. Consider the intersection $A_S = \bigcap_{i \in S} K_i \cap \bigcap_{i \in \overline{S}} \overline{K_i}$, where $S \subseteq Q_n$, and $\overline{S} = Q_n \setminus S$. Clearly $n-1$ must be in $\overline{S}$ if A_S is an atom, for an atom must be non-empty. Since a prefix-free language has only one final state and that state accepts ε, if $n-2 \in S$, no other quotient is in S, for then A_S would not be an atom. Hence if $S = \{n-2\}$ then $A_S = \{\varepsilon\}$, and $\kappa(A_S) = 2$.

   Now suppose $S = \emptyset$; then $A_S = \bigcap_{i \in \overline{S}} \overline{K_i}$. Since K_{n-1} appears in every quotient of A_S, there are at most 2^{n-1} subsets of Q_{n-1} that can be reached from A_S together with $n-1$. Hence $\kappa(A_S) \le 2^{n-1}$.

   If $S = Q_{n-2}$, then $\overline{S} = \{n-2, n-1\}$ and $\bigcap_{i \in \overline{S}} \overline{K_i} = \Sigma^+$. If we reach $\overline{K_{n-1}} = \Sigma^*$, then any intersection which has $\{n-2, n-1\}$ in the complemented part is equivalent to one that has only $\{n-1\}$, since no quotient other than K_{n-2} contains ε. Hence we can reach at most $2^{n-2}-1$ subsets of Q_{n-2}, along with the intersection $K_{n-2} \cap \overline{K_{n-1}} = \varepsilon$, and the empty quotient, for a total of $2^{n-2}+1$ states.

   Finally, consider the case where $\emptyset \subsetneq S \subsetneq Q_{n-2}$. Then we have from 1 to $|S|$ uncomplemented quotients K_i with $i \in Q_{n-2}$, and from 1 to $n-2-|S|$ quotients K_i with $i \in Q_{n-2}$ in the complemented part; this leads to the formula given in the theorem.

It remains to be proved that the atoms of $L_n(a,b,c,-,e_0)$ meet these bounds. Atom $A_{\{n-2\}}$ is equal to $\{\varepsilon\}$ and thus has two quotients as required; assume now that $S \subseteq Q_{n-2}$. We are interested in the number of distinct quotients of $A_S = \bigcap_{i\in S} K_i \cap \bigcap_{i\in \overline{S}} \overline{K_i}$, where $S \subseteq Q_n \setminus \{n-1\}$. The quotients $w^{-1}A_S$ have the form $J_{X,Y} = \bigcap_{i\in X} K_i \cap \bigcap_{i\in Y} \overline{K_i}$ where $X = \{i \mid K_i = w^{-1}K_j \text{ for some } j \in S\}$ and $Y = \{i \mid K_i = w^{-1}K_j \text{ for some } j \in \overline{S}\}$. For brevity, we write $S \xrightarrow{w} X$ and $\overline{S} \xrightarrow{w} Y$; this notation is in agreement with the action of w on the states of $\mathcal{D}_n$ corresponding to S and $\overline{S}$.

Notice $J_{X,Y} = J_{X,Y\cup\{n-2\}}$ for all X and Y, except for the case $X = \{n-2\}$ in which $J_{X,Y} \in \{\{\varepsilon\},\emptyset\}$. Thus it is sufficient to assume $n-2 \notin Y$ from now on, as $\{J_{X,Y} \mid n-2 \notin Y, n-1 \in Y\}$ contains every quotient of A_S. We show that whenever $|X| \le |S|$, $|Y| \le |\overline{S}|$, $n-2 \notin Y$, and $n-1 \in Y$, there is a word $w \in \{a,b,c,e_0\}^*$ such that $S \xrightarrow{w} X$ and $\overline{S} \xrightarrow{w} Y$ and hence $J_{X,Y}$ is a quotient of A_S. When $S = Q_{n-2}$ we reach all quotients $J_{X,\{n-1\}}$ where $\emptyset \subsetneq X \subseteq Q_{n-2}$ by words in $\{a,b,c\}^*$, we reach $J_{\{n-2\},\{n-1\}}$ from $J_{\{0\},\{n-1\}}$ by e_0, and from there we reach the empty quotient by e_0. Similarly, when $\emptyset \subseteq S \subsetneq Q_{n-2}$, we reach $J_{X,Y}$ for $X \subseteq Q_{n-2}$ and $Y \cap Q_{n-2} \ne \emptyset$ by words in $\{a,b,c\}^*$, and the remaining quotients are easily reached using e_0.

It remains to show that non-empty quotients $J_{X,Y}$ and $J_{X',Y'}$ are distinct whenever $X \ne X'$ or $Y \ne Y'$. Notice $J_{X,Y} = \emptyset$ if either $X \cap Y \ne \emptyset$ or $\{n-2\} \subsetneq X$, and $J_{X,Y} = \{\varepsilon\}$ if and only if $X = \{n-2\}$. Apart from these special cases, every $J_{X,Y}$ is non-empty and does not contain ε.

For any $X \subseteq Q_{n-2}$, let w_X denote a word that maps $X \to \{n-2\}$ and $Q_n \setminus X \to \{n-1\}$; there is such a word in $\{a,b,c,e_0\}^*$ because $\{a,b,c\}^*$ contains u: $(n{-}2 \to n{-}1)(X \to n{-}3)(Q_{n-2}\setminus X \to 0)$, and then $w_X = ue_0ae_0$. Observe that $w_X \in K_i$ for all $i \in X$ and $w_X \notin K_j$ for all $j \notin X$. Hence, if $X \subseteq Q_{n-2}$ and $Y \subseteq Q_n \setminus X$, then $w_X \in J_{X,Y}$ and $w_{\overline{Y}\cap Q_{n-2}} \in J_{X,Y}$.

Let X' and Y' be any disjoint subsets of Q_n where $n-1 \in Y'$ and $J_{X',Y'} \ne \emptyset$. If $X' \ne X$ then either $w_X \notin J_{X',Y'}$ or $w_{X'} \notin J_{X,Y}$. Similarly, if $Y' \ne Y$ (and $Y \oplus Y' \ne \{n-2\}$) then either $w_{\overline{Y}\cap Q_{n-2}} \notin J_{X',Y'}$ or $w_{\overline{Y'}\cap Q_{n-2}} \notin J_{X,Y}$. Thus, any two quotients $J_{X,Y}$ and $J_{X',Y'}$, where $(X,Y) \ne (X',Y')$, are distinct.

When we established the upper bound on $\kappa(A_S)$, we counted the number of reachable, potentially distinct quotients $J_{X,Y}$ of each A_S. We have now shown that every reachable $J_{X,Y}$ is a quotient of A_S and determined all the cases when $J_{X,Y} = J_{X',Y'}$. It follows that every bound is met by $L_n(a,b,c,-,e_0)$.

5. **Star** Proved in [19]. For the purpose of proving that $n+2$ inputs are required for a most complex prefix-free witness, an outline of the proof is repeated here.

   Suppose that L is a prefix-free language with n quotients whose syntactic semigroup is maximal, and L^* has maximal complexity. We show that L requires an alphabet of size $n+2$. Towards a contradiction, let $\mathcal{D} = (Q_n, \Sigma, \delta, 0, \{n-2\})$ be a DFA for L where $|\Sigma| = n+1$. Assume $0,1,\dots,n-3$ are non-final, non-empty states, $n-2$ is the unique final state, and $n-1$ is the empty state. By [19], $\mathcal{D}$ must have this structure and $\delta(n-2, w) = n-1$ for any $w \in \Sigma^+$.

Since the syntactic semigroup of L is maximal, each letter of Σ has a specific role in $\mathcal{D}$ as described in **1** of this theorem. Three letters a', b', and c' are required to induce the transformations on Q_{n-2}; these letters cannot map any state of Q_{n-2} to $n-2$ or to $n-1$. An additional $n-2$ letters $v_0, v_1, \dots, v_{n-3}$ are required to generate $e_q \colon (n{-}2 \to n{-}1)(q \to n{-}2)$ for each $q \in Q_{n-2}$, where the action of e_q is induced by a word in $\{a', b', c'\}^* v_q$. Notice v_q cannot map any state of Q_{n-2} to $n-1$, since e_q does not. In summary, $\Sigma = \{a', b', c', v_0, \dots, v_{n-3}\}$ and for all $\ell \in \Sigma$ and $q \in Q_{n-2}$, $\delta(q, \ell) \neq n-1$.

An NFA for L^* is produced by adding to $\mathcal{D}$ a new initial state $0'$, which is final, adding an ε-transition from $n{-}2$ to 0, and deleting the empty state $n{-}1$. The transitions from $0'$ are exactly the same as the transitions from 0. Perform the subset construction on this NFA. The $n-1$ states $\{0'\}, \{0\}, \{1\}, \dots, \{n-3\}$ are all reachable and distinguishable by words in $\{a', b', c', v_0\}$. The only way to reach a set containing more than one state is by moving to $n-2$ and using the ε-transition. This leads to the state $\{0, n-2\}$, but applying any word $w \in \Sigma^+$ deletes $n-2$; thus, $\{0, n-2\}$ is the only reachable set with two or more states. However, $\{0'\}$ and $\{0, n-2\}$ are indistinguishable, since both are final and $\delta(\{0'\}, w) = \delta(\{0\}, w) = \delta(\{0, n-2\}, w)$ for $w \in \Sigma^+$.

So far, there are only $n-1$ reachable, distinguishable states in the subset construction. The remaining state is $\emptyset$, which can only be reached if there is a letter ℓ that moves from $q \in Q_{n-2}$ to $n-1$ in $\mathcal{D}$; a transition from $n-2$ to $n-1$ is not sufficient to reach the empty state. We showed that in our witness no $\ell \in \Sigma$ has $\delta(q, \ell) = n{-}1$. Therefore, $\kappa(L^*) \le n{-}1$, a contradiction. To achieve $\kappa(L^*) = n$, an additional letter is required. Therefore, any most complex prefix-free language stream requires $n+2$ inputs.

6. **Product** Proved in [19].

7. **Boolean Operations** Let $S = Q'_{m-2} \times Q_{n-2}$. For $0 \le p \le m-1$, let $R_p = \{(p', q) \mid q \in Q_n\}$, and for $0 \le q \le n-1$ let $C_q = \{(p', q) \mid p' \in Q'_m\}$. These are the sets of states in the rows and columns of Figure 13. The states in S are reachable from the initial state $(0', 0)$ by [2, Theorem 1]. Every other state in the direct product is reachable from some state in S, as illustrated in Figure 13.

   For $\circ \in \{\cup, \oplus, \setminus, \cap\}$, the direct product recognizes $L'_m \circ L_n$ if the final states are set to be $R_{m-2} \circ C_{n-2}$. Now we must determine which states are distinguishable with respect to $R_{m-2} \circ C_{n-2}$ for each value of $\circ$. Consider the DFAs $D'_m = (Q'_{m-2}, \{a, b\}, \delta, 0', \{(m-3)'\})$ and $D_n = (Q_{n-2}, \{b, a\}, \delta, 0, \{n-3\})$. By [2, Theorem 1], the states of S are pairwise distinguishable with respect to $(R_{m-3} \circ C_{n-3}) \cap S$. For any pair of states in S, let w be a word that distinguishes them in $(R_{m-3} \circ C_{n-3}) \cap S$; one verifies that further apply-

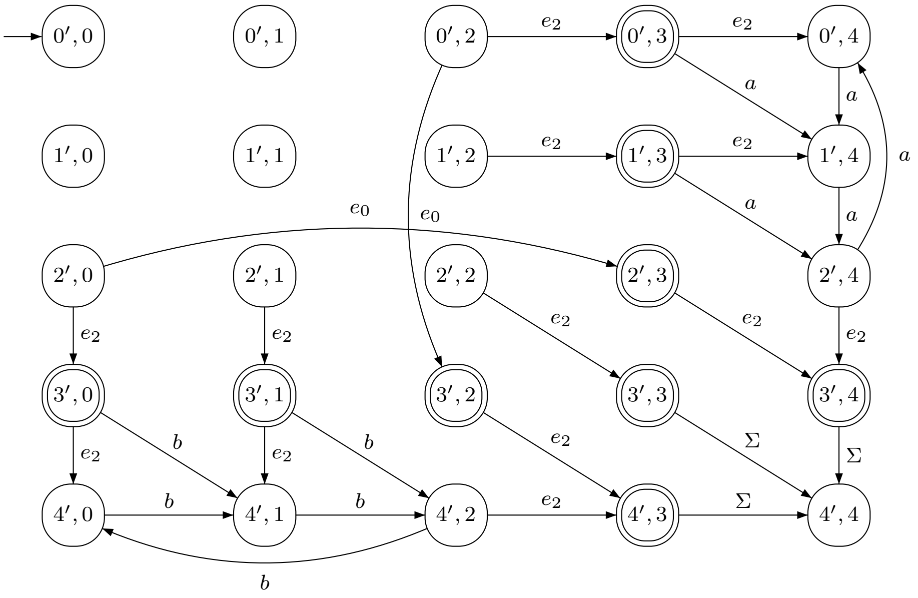

Figure 13: Partial illustration of the direct product for $L'_5(a, b, -, -, e_0, e_2) \cup L_5(b, a, -, -, e_0, e_2)$.

ing e_{m-3} distinguishes them with respect to $R_{m-2} \circ C_{n-2}$. The rest of the distinguishability argument depends on $\circ \in \{\cup, \oplus, \setminus, \cap\}$.

$\cup$: States $((m-1)', n-2)$, $((m-2)', n-1)$, and $((m-2)', n-2)$ are equivalent, since all three are final and any letter sends them to $((m-1)', n-1)$.

States of R_{m-1} are distinguished by words in $b^* e_{m-3}$. States of C_{n-1} are distinguished by words in $a^* e_{m-3}$. Excluding $((m-1)', n-2)$ and $((m-2)', n-1)$, which are equivalent, states of R_{m-1} are distinguished from states of C_{n-1} by words in $a^* e_{m-3}$.

States of $R_{m-2} \cup C_{n-2}$ are moved to states of $R_{m-1} \cup C_{n-1}$ by applying e_{m-3}. Excluding $((m-1)', n-2)$, $((m-2)', n-1)$, and $((m-2)', n-2)$, which are equivalent, every state is mapped by e_{m-3} to a different state of $R_{m-1} \cup C_{n-1}$; hence they are distinguishable.

Finally, we must show that states of S are distinguishable from the states of $R_{m-1} \cup C_{n-1}$. For any $(p', q) \in S$, there exists $w \in \{a, b\}^*$ such that $(p', q) \xrightarrow{w} (0', n-3)$, since both (p', q) and $(0', n-3)$ are reached from $(0', 0)$ by words in $\{a, b\}^*$, and a and b permute S. Then $(0', n-3) \xrightarrow{e_{m-3}} (0', n-2)$ and we have already shown that $(0', n-2)$ is distinguishable from all states in $R_{m-1} \cup C_{n-1}$. Thus, the $mn-2$ remaining states are pairwise distinguishable.

$\oplus$: States $((m-1)', n-2)$ and $((m-2)', n-1)$ are equivalent, and states

$((m-2)', n-2)$ and $((m-1)', n-1)$ are equivalent. The rest of the states are distinguishable by an argument similar to that of union.

$\cap$: State $((m-2)', n-2)$ is the only final state. The remaining non-final states of $R_{m-2} \cup R_{m-1} \cup C_{n-2} \cup C_{n-1}$ are all empty. Clearly the states of S are non-empty, since $((m-3)', n-3) \xrightarrow{e_{m-3}} ((m-2)', n-2)$. Thus, the remaining $mn - 2(m+n-3)$ states are pairwise distinguishable.

$\setminus$: The states of R_{m-1} and $((m-2)', n-2)$ are all equivalent. States $((m-1)', q)$ and $((m-2)', q)$ are equivalent for $0 \le q \le m-3$. The final states $(R_{m-2} \setminus \{((m-2)', n-2)\})$ are all equivalent.

The states of C_{n-1} are distinguished by words in a^*e_{m-3}. It remains to show that states of S are distinguishable from the states of C_{n-1}. Notice $((m-3)', n-3)$ is distinguished from $((m-3)', n-1)$ by e_{m-3}, and from every other state of C_{n-1} by be_{m-3}. For any state of S, there exists $w \in \{a,b\}^*$ that sends that state to $((m-3)', n-3)$, and notice $C_{n-1}w \subseteq C_{n-1}$. So all $mn - (m + 2n - 4)$ remaining states are pairwise distinguishable.

Note that stream $L_m(a, b, c, d, e_0, e_{m-3})$ with dialect $L_n(b, a, c, d, e_0, e_{m-3})$ meets the bounds for quotients, reversal, atomic complexity, star, product and boolean operations. □

Using some results from [22, 23] we define another prefix-free witness stream that meets all the bounds except those for syntactic complexity and atom complexity. *Moreover, all the bounds are met by dialects over minimal alphabets.*

Definition 5. *For $n \ge 4$, let $\mathcal{D}_n(a, c, d, e, f, g) = (Q_n, \Sigma, \delta_n, 0, \{n-2\})$, where $\Sigma = \{a, c, d, e, f, g\}$, and δ_n is defined by the transformations*

- $a\colon (n-2 \to n-1)(0, \dots, n-3)$,
- $c\colon (n-2 \to n-1)(1 \to 0)$.
- $d\colon (0 \to n-2)(Q_n \setminus \{0\} \to n-1)$,
- $e\colon (n-2 \to n-1)(n-3 \to n-2)$,
- $f\colon (n-2 \to n-1)({}_0^{n-2}\ q \to q+1)$,
- $g\colon (n-2 \to n-1)$.

Note that b is not used, a, c, d, and e induce the same transformations as a, c, d, and e_{n-3} in Definition 4. DFA $\mathcal{D}_n(\Sigma)$ is shown in Figure 14. Let $L_n(\Sigma)$ be the language accepted by $\mathcal{D}_n(\Sigma)$.

Proposition 1. *For $n \ge 4$, the DFA of Definition 5 is minimal and $L_n(\Sigma)$ is a prefix-free language of complexity n. Moreover, all the witnesses for individual operations have minimal alphabets.*

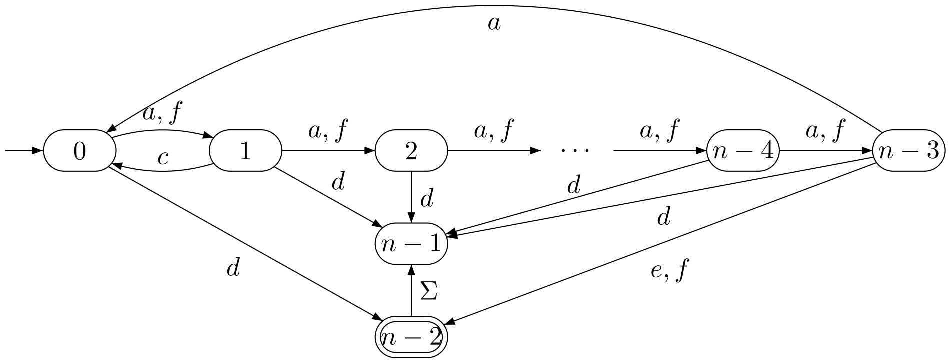

Figure 14: DFA $\mathcal{D}_n(\Sigma)$ of Definition 5; missing transitions are self-loops.

1. *The quotients of* $L_n(a,-,-,-,f)$ *have complexity* n*, except for the quotient* ε *and the empty quotient, which have complexity 2 and 1 respectively.*
2. *The reverse of* $L_n(a,c,-,e)$ *has complexity* $2^{n-2}+1$*, and* $L_n(a,-,c,d)$ *has* $2^{n-2}+1$ *atoms.*
3. *The star of* $L_n(a,-,-,d)$ *has complexity* n*.*
4. *For* $m,n \geq 4$*,* $\kappa(L_m(-,-,-,-,f)L_n(-,-,-,-,f)) = m+n-2$*.*
5. *a)* $\kappa(L_m(-,-,-,-,f,g) \cup L_n(-,-,-,-,g,f)) = \kappa(L_m(-,-,-,-,f,g) \oplus L_n(-,-,-,-,g,f)) = mn-2$*.*

   b) $\kappa(L_m(a,-,-,e,-,-) \setminus L_n(-,-,-,-,e,a)) = mn-(m+2n-4)$*.*

   c) $\kappa(L_m(a,-,-,e,-,-) \cap L_n(-,-,-,-,e,a)) = mn-2(m+n-3)$*.*

Proof. The first claim is obvious. The second and third claims were proved in Theorem 4. (A ternary witness was also used in [22] for the reverse, but it had more complicated transitions than our witness.) The fourth claim is from [22]. The results for union, symmetric difference and intersection were proved in [22], and that for difference in [23]. □

8 Conclusions

Our results are summarized in Table 1. The largest bounds are shown in boldface type, and they are reached in the classes of ideal and closed languages. Recall that for regular languages we have the following results: semigroup: n^n; reverse: 2^n; star: $2^{n-1}+2^{n-2}$; restricted product: $(m-1)2^n+2^{n-1}$; unrestricted product: $m2^n+2^{n-1}$; restricted $\cup$ and $\oplus$: mn; unrestricted $\cup$ and $\oplus$: $(m+1)(n+1)$; restricted $\setminus$: mn; unrestricted $\setminus$: $mn+m$; restricted $\cap$: mn; unrestricted $\cap$: mn.

Table 1: Complexities of special prefix-convex languages

	Right-Ideal	Prefix-Closed	Prefix-Free
Semigroup	$\mathbf{n^{n-1}}$	$\mathbf{n^{n-1}}$	n^{n-2}
Reverse	$\mathbf{2^{n-1}}$	$\mathbf{2^{n-1}}$	$2^{n-2}+1$
Star	$n+1$	$\mathbf{2^{n-2}+1}$	n
Product restricted	$m+2^{n-2}$	$\mathbf{(m+1)2^{n-2}}$	$m+n-2$
Product unrestr.	$m+2^{n-1}+2^{n-2}+1$	$\mathbf{(m+1)2^{n-2}}$	$m+n-2$
$\cup$ restricted	$mn-(m+n-2)$	$\mathbf{mn}$	$mn-2$
$\cup$ unrestricted	$\mathbf{(m+1)(n+1)}$	mn	$mn-2$
$\oplus$ restricted	$\mathbf{mn}$	$\mathbf{mn}$	$mn-2$
$\oplus$ unrestricted	$\mathbf{(m+1)(n+1)}$	mn	$mn-2$
$\setminus$ restricted	$\mathbf{mn-(m-1)}$	$\mathbf{mn-(n-1)}$	$mn-(m+2n-4)$
$\setminus$ unrestricted	$\mathbf{mn+m}$	$\mathbf{mn-(n-1)}$	$mn-(m+2n-4)$
$\cap$ restr. and unrestr.	$\mathbf{mn}$	$mn-(m+n-2)$	$mn-2(m+n-3)$

References

[1] Ang, T. and Brzozowski, J. A. Languages convex with respect to binary relations, and their closure properties. *Acta Cybernet.*, 19(2):445–464, 2009.

[2] Bell, J., Brzozowski, J. A., Moreira, N., and Reis, R. Symmetric groups and quotient complexity of boolean operations. In Esparza, J. and et al., editors, *ICALP 2014*, volume 8573 of *LNCS*, pages 1–12. Springer Berlin / Heidelberg, 2014.

[3] Berstel, J., Perrin, D., and Reutenauer, C. *Codes and Automata (Encyclopedia of Mathematics and its Applications).* Cambridge University Press, 2010.

[4] Brzozowski, J. A. Quotient complexity of regular languages. *J. Autom. Lang. Comb.*, 15(1/2):71–89, 2010.

[5] Brzozowski, J. A. In search of the most complex regular languages. *Int. J. Found. Comput. Sci,*, 24(6):691–708, 2013.

[6] Brzozowski, J. A. Unrestricted state complexity of binary operations on regular languages. In C. Câmpeanu, F. Manea and Shallit, J., editors, *DCFS 2016*, volume 9777 of *LNCS*, pages 60–72. Springer Berlin / Heidelberg, 2016.

[7] Brzozowski, J. A. and Davies, G. Maximally atomic languages. In Ësik, Z. and Fülop, Z., editors, *Automata and Formal Languages (AFL 2014)*, pages 151–161. EPTCS, 2014.

[8] Brzozowski, J. A. and Davies, S. Quotient complexities of atoms in regular ideal languages. *Acta Cybernet.*, 22(2):293–311, 2015.

[9] Brzozowski, J. A., Davies, S., and Liu, B. Y. V. Most complex regular ideal languages. *Discrete Math. Theoret. Comput. Sc.*, 18(3), 2016. Paper #15.

[10] Brzozowski, J. A., Jirásková, G., and Li, B. Quotient complexity of ideal languages. *Theoret. Comput. Sci.*, 470:36–52, 2013.

[11] Brzozowski, J. A., Jirásková, G., and Zou, C. Quotient complexity of closed languages. *Theory Comput. Syst.*, 54:277–292, 2014.

[12] Brzozowski, J. A., Li, B., and Ye, Y. Syntactic complexity of prefix-, suffix-, bifix-, and factor-free regular languages. *Theoret. Comput. Sci.*, 449:37–53, 2012.

[13] Brzozowski, J. A. and Liu, B. Quotient complexity of star-free languages. *Int. J. Found. Comput. Sci.*, 23(6):1261–1276, 2012.

[14] Brzozowski, J. A. and Sinnamon, C. Unrestricted state complexity of binary operations on regular and ideal languages. *J. Autom. Lang. Comb.* To appear. See also `http://arxiv.org/abs/1609.04439`.

[15] Brzozowski, J. A., Szykuła, M., and Ye, Y. Syntactic complexity of regular ideals. `http://arxiv.org/abs/1509.06032`.

[16] Brzozowski, J. A. and Tamm, H. Quotient complexities of atoms of regular languages. *Int. J. Found. Comput. Sci.*, 24(7):1009–1027, 2013.

[17] Brzozowski, J. A. and Tamm, H. Theory of átomata. *Theoret. Comput. Sci.*, 539:13–27, 2014.

[18] Brzozowski, J. A. and Ye, Y. Syntactic complexity of ideal and closed languages. In Mauri, G. and Leporati, A., editors, *DLT 2011*, volume 6795 of *LNCS*, pages 117–128. Springer Berlin / Heidelberg, 2011.

[19] Han, Y.-S., Salomaa, K., and Wood, D. Operational state complexity of prefix-free regular languages. In Ésik, Z. and Fülöp, Z, editors, *Automata, Formal Languages, and Related Topics*, pages 99–115. Institute of Informatics, University of Szeged, Hungary, 2009.

[20] Holzer, M. and König, B. On deterministic finite automata and syntactic monoid size. *Theoret. Comput. Sci.*, 327(3):319–347, 2004.

[21] Iván, S. Complexity of atoms, combinatorially. *Inform. Process. Lett.*, 116(5):356–360, 2016.

[22] Jirásková, G. and Krausová, M. Complexity in prefix-free regular languages. In McQuillan, I., Pighizzini, G., and Trost, B., editors, *Proceedings of the 12th International Workshop on Descriptional Complexity of Formal Systems (DCFS)*, pages 236–244. University of Saskatchewan, 2010.

[23] Krausová, M. Prefix-free regular languages: Closure properties, difference, and left quotient. In Kotásek, Z., Bouda, J., Cerná, I., Sekanina, L., Vojnar, T., and Antos, D., editors, *MEMICS*, volume 7119 of *Lecture Notes in Computer Science*, pages 114–122. Springer Berlin / Heidelberg, 2011.

[24] Krawetz, B., Lawrence, J., and Shallit, J. State complexity and the monoid of transformations of a finite set. In Domaratzki, M., Okhotin, A., Salomaa, K., and Yu, S., editors, *Proceedings of the Implementation and Application of Automata, (CIAA)*, volume 3317 of *LNCS*, pages 213–224. Springer Berlin / Heidelberg, 2005.

[25] Myhill, J. Finite automata and representation of events. *Wright Air Development Center Technical Report*, 57–624, 1957.

[26] Pin, J.-E. Syntactic semigroups. In *Handbook of Formal Languages, vol. 1: Word, Language, Grammar*, pages 679–746. Springer, New York, NY, USA, 1997.

[27] Salomaa, A., Wood, D., and Yu, S. On the state complexity of reversals of regular languages. *Theoret. Comput. Sci.*, 320:315–329, 2004.

[28] Thierrin, G. Convex languages. In Nivat, M., editor, *Automata, Languages and Programming*, pages 481–492. North-Holland, 1973.

[29] Yu, S. State complexity of regular languages. *J. Autom. Lang. Comb.*, 6:221–234, 2001.