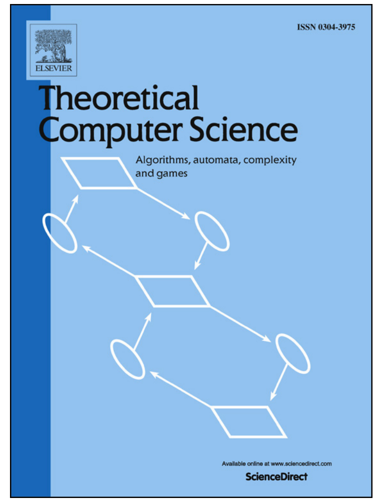

Complexity of proper prefix-convex regular languages

Janusz A. Brzozowski, Corwin Sinnamon

PII: S0304-3975(18)30496-1
DOI: https://doi.org/10.1016/j.tcs.2018.07.015
Reference: TCS 11691

To appear in: *Theoretical Computer Science*

Received date: 24 December 2017
Revised date: 27 June 2018
Accepted date: 17 July 2018

Please cite this article in press as: J.A. Brzozowski, C. Sinnamon, Complexity of proper prefix-convex regular languages, *Theoret. Comput. Sci.* (2018), https://doi.org/10.1016/j.tcs.2018.07.015

Complexity of Proper Prefix-Convex Regular Languages☆

Janusz A. Brzozowski[a], Corwin Sinnamon[a,*]

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

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, which were studied elsewhere. Here we concentrate on prefix-convex languages that do not belong to any one of these classes; we call such languages *proper*. We exhibit most complex proper prefix-convex languages, which meet the bounds for the size of the syntactic semigroup, reversal, complexity of atoms, star, product, and boolean operations.

Keywords: atom, most complex, prefix-convex, proper, quotient complexity, regular language, state complexity, syntactic semigroup

1. Introduction

Prefix-Convex Languages We examine the complexity properties of a class of regular languages that has never been studied before: the class of proper prefix-convex languages [7]. Let Σ be a finite alphabet; if $w = xy$, for $x, y \in \Sigma^*$, then x is a prefix of w. A language $L \subseteq \Sigma^*$ is *prefix-convex* [1, 17] if whenever x and xyz are in L, then so is xy. Prefix-convex languages include three special cases:

1. A language $L \subseteq \Sigma$ is a *right ideal* if it is non-empty and satisfies $L = L\Sigma^*$. Right ideals appear in pattern matching [11]: $L\Sigma^*$ is the set of all words in some text (word in Σ^*) beginning with words in L.
2. A language is *prefix-closed* [6] if whenever w is in L, then so is every prefix of w. The set of allowed sequences to any system is prefix-closed. Every prefix-closed language other than Σ^* is the complement of a right ideal [1].

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

*Corresponding author
Email addresses: brzozo@uwaterloo.ca (Janusz A. Brzozowski), sinncore@gmail.com (Corwin Sinnamon)

3. A language is *prefix-free* if $w \in L$ implies that no prefix of w other than w is in L. Prefix-free languages other than $\{\varepsilon\}$, where ε is the empty word, are prefix codes and are of considerable importance in coding theory [2].

The complexities of these three special prefix-convex languages were studied in [8]. We now turn to the "real" prefix-convex languages that do not belong to any of the three special classes.

Complexities of Operations If $L \subseteq \Sigma^*$ is a language, 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, the *quotient complexity* [3] $\kappa(L)$ of L, is a natural measure of complexity for L. An equivalent concept is the *state complexity* [12, 16, 18, 19] of L, which is the number of states in a complete minimal deterministic finite automaton (DFA) over Σ recognizing L. We refer to quotient/state complexity simply as *complexity*.

If L_n is a regular language of complexity n, and $\circ$ is a unary operation, the *complexity of* $\circ$ is the maximal value of $\kappa(L_n^\circ)$, expressed as a function of n, as L_n ranges over all languages of complexity $\leqslant n$. If L'_m and L_n are regular languages[1] of complexities m and n respectively, and $\circ$ is a binary operation, the *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 languages of complexities $\leqslant m$ and $\leqslant n$. The complexity of an operation is a lower bound on its time and space complexities. The operations reversal, (Kleene) star, product (concatenation), and binary boolean operations are considered "common", and their complexities are known; see [4, 12, 18, 19].

Witnesses To find the complexity of a unary operation we find an upper bound on this complexity, and languages that meet this bound. We require a language L_n for each n, that is, a sequence, $(L_k, L_{k+1}, \dots)$, called a *stream* of languages. A stream begins at k, a small integer, because the bound may not hold for small values of n. For a binary operation we need two streams. The same stream cannot always be used for both operands, but for all common binary operations the second stream can be a "dialect" of the first, that is it can "differ only slightly" from the first [4]. Let $\Sigma = \{a_1, \dots, a_k\}$ be an alphabet ordered as shown; if $L \subseteq \Sigma^*$, we denote it by $L(a_1, \dots, a_k)$. A *dialect* of L is obtained by deleting letters of Σ in the words of L, or replacing them by letters of another alphabet Σ'. More precisely, for an injective partial map $\pi\colon \Sigma \mapsto \Sigma'$, we get a dialect of L by replacing each letter $a \in \Sigma$ by $\pi(a)$ in every word of L, or deleting the word 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 "$-$". Undefined values for letters at the end of the alphabet are omitted; for example, $L(a, c, -, -)$ is written as $L(a, c)$. Our definition of dialect is more general than that of [5], where only the case $\Sigma' = \Sigma$ was allowed.

[1]We often use the variable names L'_m and L_n when two different languages are needed. The primed variable does not have any special meaning.

Finite Automata 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 set of all words accepted by $\mathcal{D}$ is the *language* $L(\mathcal{D})$ *of* $\mathcal{D}$. If $q \in Q$, then the *language* $L_q(D)$ *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 $L_p(\mathcal{D}) = L_q(\mathcal{D})$. 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. A *nondeterministic finite automaton (NFA)* is a quintuple $\mathcal{D} = (Q, \Sigma, \delta, I, F)$, where Q, Σ, and F are defined as in a DFA, $\delta\colon Q \times \Sigma \to 2^Q$ is the *transition function*, 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.

Transformations 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 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 $a\colon \delta_a$. Often 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: if $P \subseteq Q_n$, then $Pa = \{pa \mid p \in P\}$. We also write $P \xrightarrow{a} Pa$ to indicate that the image of P under a is Pa. If s, t are transformations of Q_n, their composition is $(qs)t$.

For $k \geqslant 2$, a transformation (permutation) t of a set $P = \{q_0, q_1, \dots, q_{k-1}\} \subseteq Q_n$ is a k*-cycle* if $q_0t = q_1, q_1t = q_2, \dots, q_{k-2}t = q_{k-1}, q_{k-1}t = q_0$. As a transformation of Q_n, this k-cycle is denoted by $(q_0, q_1, \dots, q_{k-1})$, and leaves the states in $Q_n \setminus P$ unchanged. 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 other states is denoted by $(P \to q)$, and $(Q_n \to p)$ is called a *constant* transformation. If $P = \{p\}$ we write $(p \to q)$ for $(\{p\} \to q)$. The identity transformation is denoted by $\mathbb{1}$. Also, $(_i^j\ q \to q+1)$ is a transformation that sends q to $q+1$ for $i \leqslant q \leqslant j$ and is the identity for the remaining states; $(_i^j\ q \to q-1)$ is defined similarly.

Semigroups The *syntactic congruence* of $L \subseteq \Sigma^*$ is defined on Σ^+: For $x, y \in \Sigma^+$, $x \approx_L y$ if and only if $wxz \in L \Leftrightarrow wyz \in L$ for all $w, z \in \Sigma^*$. The quotient set $\Sigma^+/{\approx_L}$ of equivalence classes of $\approx_L$ is the *syntactic semigroup* of L. Let $\mathcal{D}_n = (Q_n, \Sigma, \delta, q_0, F)$ be a DFA, and let $L_n = L(\mathcal{D}_n)$. 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 $T_{\mathcal{D}_n}$ of all such transformations by non-empty words is a semigroup under composition called the *transition semigroup* of $\mathcal{D}_n$. If $\mathcal{D}_n$ is a minimal DFA of L_n, then $T_{\mathcal{D}_n}$ is isomorphic to the syntactic semigroup T_{L_n} of L_n, and we represent elements of T_{L_n} by transformations in $T_{\mathcal{D}_n}$. The size of the syntactic semigroup has been used as a measure of complexity for regular languages [4, 10, 13, 15].

Atoms are defined by a left congruence, where 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 [9, 14].

One can conclude that an atom is a non-empty intersection of complemented and uncomplemented quotients of L. That is, every atom of a language with quotients $K_0, K_1, \dots, K_{n-1}$ can be written as $A_S = \bigcap_{i \in S} K_i \cap \bigcap_{i \in \overline{S}} \overline{K_i}$ for some set $S \subseteq Q_n$. The number of atoms and their complexities were suggested as possible measures of complexity [4], because all the quotients of a language and the quotients of its atoms are unions of atoms [9].

Most Complex Regular Stream The stream $(\mathcal{D}_n(a,b,c) \mid n \geqslant 3)$ of Definition 1 and Figure 1 will be used as a component in the class of proper prefix-convex languages. This stream together with some dialects meets the complexity bounds for reversal, star, product, and all binary boolean operations [7, 8]. Moreover, it has the maximal syntactic semigroup and most complex atoms, making it a most complex regular stream.

Definition 1. For $n \geqslant 3$, 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 $a\colon (0,\dots,n-1)$, $b\colon (0,1)$, $c\colon (1 \to 0)$.

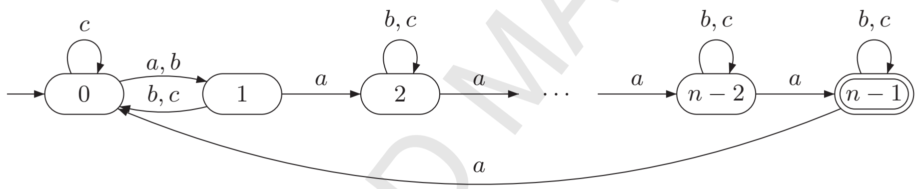

Figure 1: Minimal DFA of a most complex regular language.

Most complex streams are useful in systems dealing with regular languages and finite automata. To know the maximal sizes of automata that can be handled by a system it suffices to use the most complex stream to test all the operations.

2. Proper Prefix-Convex Languages

We begin with some properties of prefix-convex languages that will be used frequently in this section. The following lemma and propositions characterize the classes of prefix-convex languages in terms of their minimal DFAs.

Lemma 1. *Let L be a prefix-convex language over Σ. Either L is a right ideal or L has an empty quotient.*

Proof. Suppose that L is not a right ideal. If $L = \emptyset$, then $\varepsilon^{-1}L = L$ is an empty quotient of L. If $L \neq \emptyset$, we cannot have $w^{-1}L = \Sigma^*$ for all $w \in L$, because then L would be a right ideal. Hence there exists some $w \in L$ such that $w^{-1}L \neq \Sigma^*$. Pick any $x \in \Sigma^* \setminus w^{-1}L$; then $w \in L$, but $wx \notin L$. There cannot be a word $y \in \Sigma^*$ such that $wxy \in L$ because then wx would be in L by prefix convexity. Therefore, $(wx)^{-1}L$ is an empty quotient. $\square$

Proposition 2. *Let L_n be a regular language of complexity n, and let $\mathcal{D}_n = (Q_n, \Sigma, \delta, 0, F)$ be a minimal DFA recognizing L_n. The following are equivalent:*

1. *L_n is prefix-convex.*
2. *For all $p, q, r \in Q_n$, if p and r are final, q is reachable from p, and r is reachable from q, then q is final.*
3. *Every state reachable in $\mathcal{D}_n$ from any final state is either final or empty.*

Proof. (**1** $\implies$ **2**) Assume **1** is true. Suppose there exist $p, r \in F$ and $q \in Q_n$ such that q is reachable from p and r is reachable from q. Let $w, x, y \in \Sigma^*$ be such that $0 \xrightarrow{w} p$, $p \xrightarrow{x} q$, and $q \xrightarrow{y} r$. It follows that w and wxy are both in L_n, and thus wx is in L_n by prefix convexity. Since $\delta(0, wx) = q$, state q is final.
(**2** $\implies$ **3**) Assume **2** is true. Take any $p \in F$, $q \in Q_n$, and $x \in \Sigma^*$ such that $\delta(p, x) = q$. If a final state r is reachable from q, then q is final by **2**. Otherwise, q is the empty state.
(**3** $\implies$ **1**) Assume **3** is true. Let w, x, and y be words in Σ^* such that $w \in L_n$ and $wxy \in L_n$. There are states p, q, and r in Q_n such that $\delta(0, w) = p \in F$, $\delta(0, wx) = q$, and $\delta(0, wxy) = r \in F$. State q cannot be empty because the final state r is reachable from q. Since q is reachable from final state p, it follows from **3** that q is final. Thus, $wx \in L_n$. Therefore L_n is prefix-convex. □

Proposition 3. *Let L_n be a non-empty prefix-convex language of complexity n, and let $\mathcal{D}_n = (Q_n, \Sigma, \delta, 0, F)$ be a minimal DFA recognizing L_n.*

1. *L_n is prefix-closed if and only if $0 \in F$.*
2. *L_n is prefix-free if and only if $\mathcal{D}_n$ has a unique final state p and an empty state p' such that $\delta(p, a) = p'$ for all $a \in \Sigma$.*
3. *L_n is a right ideal if and only if $\mathcal{D}_n$ has a unique final state p and $\delta(p, a) = p$ for all $a \in \Sigma$.*

Proof.

1. If L_n is prefix-closed and non-empty, then ε is a prefix of some word in L_n. Thus $\varepsilon \in L_n$, and so $0 \in F$. Conversely, suppose $0 \in F$. For any $wx \in L_n$, there are states $q, r \in Q_n$ such that $0 \xrightarrow{w} q \xrightarrow{x} r$, and r is final. By Proposition 2, since $0, r \in F$, q is reachable from 0, and r is reachable from q, we have $q \in F$. Hence $w \in L_n$, and therefore L_n is prefix-closed.
2. Suppose L_n is prefix-free. If $q \in Q_n$ and $p \in F$ are distinct and q is reachable from p, then q cannot be final as that would imply $p \notin F$. In particular, for any $p \in F$ and $a \in \Sigma$, $\delta(p, a) \notin F$. By Proposition 2, $\delta(p, a)$ must be the empty state for all $a \in \Sigma$. Thus, the transitions from all final states are identical, and hence all final states are equivalent. By minimality, $\mathcal{D}_n$ has a unique final state p, an empty state p', and $\delta(p, a) = p'$ for all $a \in \Sigma$.

For the converse, suppose $F = \{p\}$, $p' \in Q_n$ is an empty state, and $\delta(p, a) = p'$ for all $a \in \Sigma$. Then $w \in L_n$ if and only if $\delta(0, w) = p$. For all $w \in L_n$ and $a \in \Sigma$, we have $\delta(0, wa) = p'$. Thus, whenever $w \in L_n$ and $wx \in L_n$, we have $x = \varepsilon$. Therefore, L_n is prefix-free.

3. Suppose L_n is a right ideal. For all $w \in L_n$ we have $L_n \supseteq w\Sigma^*$, and hence $w^{-1}L_n \supseteq \Sigma^*$, meaning that $w^{-1}L_n = \Sigma^*$. Hence, for any final state $q \in F$ and $x \in \Sigma^*$, $\delta(q, x) \in F$. This implies that all final states are equivalent. By minimality, there is a unique final state p. Since $\delta(p, a) \in F$ for all $a \in \Sigma$, it follows that $\delta(p, a) = p$ for all $a \in \Sigma$. For the converse, suppose $F = \{p\}$ and $\delta(p, a) = p$ for all $a \in \Sigma$. Then $w \in L_n$ if and only if $\delta(0, w) = p$. Hence, for all $w \in L_n$ and $x \in \Sigma^*$, we have $\delta(0, wx) = p$. Thus, $w\Sigma^* \subseteq L_n$ for all $w \in L_n$, and so $L_n = L_n\Sigma^*$. Therefore, L_n is a right ideal. □

A prefix-convex language L is *proper* if it is not a right ideal and it is neither prefix-closed nor prefix-free. We say it is k*-proper* if it has k final states, $1 \leqslant k \leqslant n-2$. Every minimal DFA for a k-proper language with complexity n has the same general structure: there are $n-1-k$ non-final, non-empty states, k final states, and one empty state. Every letter fixes the empty state and, by Proposition 2, no letter sends a final state to a non-final, non-empty state.

Next we define a stream of k-proper DFAs and languages, which we will show to be most complex.

Definition 2. For $n \geqslant 3$, $1 \leqslant k \leqslant n-2$, let $\mathcal{D}_{n,k}(\Sigma) = (Q_n, \Sigma, \delta_{n,k}, 0, F_{n,k})$ where $\Sigma = \{a, b, c_1, c_2, d_1, d_2, e\}$, $F_{n,k} = \{n-1-k, \dots, n-2\}$, and $\delta_{n,k}$ is given by the transformations

$$a\colon \begin{cases} (1,\dots,n-2-k)(n-1-k,n-k), & \textit{if } n-1-k \textit{ is even and } k \geqslant 2;\\ (0,\dots,n-2-k)(n-1-k,n-k), & \textit{if } n-1-k \textit{ is odd and } k \geqslant 2;\\ (1,\dots,n-2-k), & \textit{if } n-1-k \textit{ is even and } k = 1;\\ (0,\dots,n-2-k), & \textit{if } n-1-k \textit{ is odd and } k = 1.\end{cases}$$

$$b\colon \begin{cases} (n-k,\dots,n-2)(0,1), & \textit{if } k \textit{ is even and } n-1-k \geqslant 2;\\ (n-1-k,\dots,n-2)(0,1), & \textit{if } k \textit{ is odd and } n-1-k \geqslant 2;\\ (n-k,\dots,n-2), & \textit{if } k \textit{ is even and } n-1-k = 1;\\ (n-1-k,\dots,n-2), & \textit{if } k \textit{ is odd and } n-1-k = 1.\end{cases}$$

$$c_1\colon \begin{cases} (1 \to 0), & \textit{if } n-1-k \geqslant 2;\\ \mathbb{1}, & \textit{if } n-1-k = 1.\end{cases}$$

$$c_2\colon \begin{cases} (n-k \to n-1-k), & \textit{if } k \geqslant 2;\\ \mathbb{1}, & \textit{if } k = 1.\end{cases}$$

$d_1\colon (n-2-k \to n-1)\binom{n-3-k}{0}\; q \to q+1)$.

$d_2\colon \binom{n-2}{n-1-k}\; q \to q+1)$.

$e\colon (0 \to n-1-k)$.

Also, let $E_{n,k} = \{0, \dots, n-2-k\}$; it is useful to partition Q_n into $E_{n,k}$, $F_{n,k}$, and $\{n-1\}$. Letters a and b have complementary behaviours on $E_{n,k}$ and $F_{n,k}$, depending on the parities of n and k. Letters c_1 and d_1 act on $E_{n,k}$ in exactly the same way as c_2 and d_2 act on $F_{n,k}$. In addition, d_1 and d_2 send states $n-2-k$ and $n-2$, respectively, to state $n-1$, and letter e connects the two parts of the DFA. The structure of $\mathcal{D}_n(\Sigma)$ is shown in Figures 2 and 3 for certain parities of $n-1-k$ and k. Let $L_{n,k}(\Sigma)$ be the language recognized by $\mathcal{D}_{n,k}(\Sigma)$.

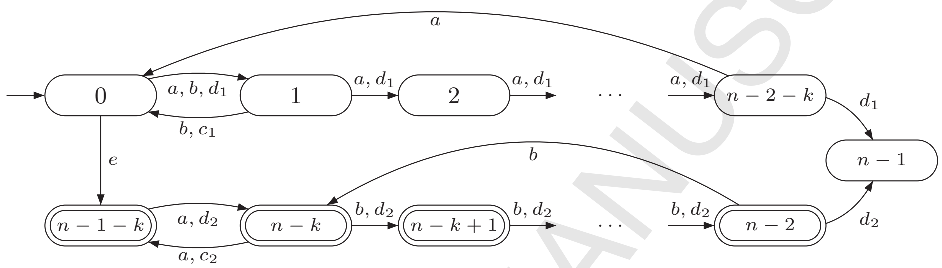

Figure 2: DFA $\mathcal{D}_{n,k}(a, b, c_1, c_2, d_1, d_2, e)$ of Definition 2 when $n-1-k$ is odd, k is even, and both are at least 2; missing transitions are self-loops.

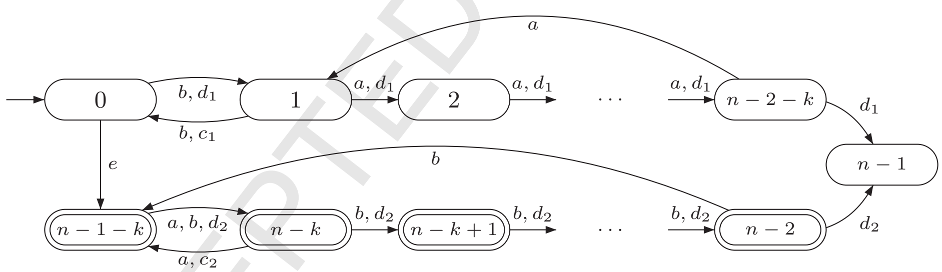

Figure 3: DFA $\mathcal{D}_{n,k}(a, b, c_1, c_2, d_1, d_2, e)$ of Definition 2 when $n-1-k$ is even, k is odd, and both are at least 2; missing transitions are self-loops.

Theorem 4 (Proper Prefix-Convex Languages). *For $n \geqslant 3$ and $1 \leqslant k \leqslant n-2$, the DFA $\mathcal{D}_{n,k}(\Sigma)$ of Definition 2 is minimal and $L_{n,k}(\Sigma)$ is a k-proper language of complexity n. The bounds below are maximal for k-proper prefix-convex languages. At least seven letters are required to meet these bounds.*

1. *The syntactic semigroup of $L_{n,k}(\Sigma)$ has cardinality $n^{n-1-k}(k+1)^k$; the maximal value $n(n-1)^{n-2}$ is reached only when $k = n-2$.*

2. *The non-empty, non-final quotients of $L_{n,k}(a, b, -, -, -, d_2, e)$ have complexity n, the final quotients have complexity $k+1$, and $\emptyset$ has complexity 1.*

3. *The reverse of* $L_{n,k}(a,b,-,-,-,d_2,e)$ *has complexity* 2^{n-1}*; moreover, the language* $L_{n,k}(a,b,-,-,-,d_2,e)$ *has* 2^{n-1} *atoms for all* k*.*

4. *For each atom* A_S *of* $L_{n,k}(\Sigma)$*, write* $S = X_1 \cup X_2$*, where* $X_1 \subseteq E_{n,k}$ *and* $X_2 \subseteq F_{n,k}$*. Let* $\overline{X_1} = E_{n,k} \setminus X_1$ *and* $\overline{X_2} = F_{n,k} \setminus X_2$*. If* $X_2 \neq \emptyset$*, then* $\kappa(A_S) =$

$$1 + \sum_{x_1=0}^{|X_1|} \sum_{x_2=1}^{|X_1|+|X_2|-x_1} \sum_{y_1=0}^{|\overline{X_1}|} \sum_{y_2=0}^{|\overline{X_1}|+|\overline{X_2}|-y_1} \binom{n-1-k}{x_1}\binom{k}{x_2}\binom{n-1-k-x_1}{y_1}\binom{k-x_2}{y_2}.$$

If $X_1 \neq \emptyset$ *and* $X_2 = \emptyset$*, then* $\kappa(A_S) =$

$$1+\sum_{x_1=0}^{|X_1|} \sum_{x_2=0}^{|X_1|-x_1} \sum_{y_1=0}^{|\overline{X_1}|} \sum_{y_2=0}^{k} \binom{n-1-k}{x_1}\binom{k}{x_2}\binom{n-1-k-x_1}{y_1}\binom{k-x_2}{y_2} - 2^k \sum_{y=0}^{|\overline{X_1}|} \binom{n-1-k}{y}.$$

Otherwise, $S = \emptyset$ *and* $\kappa(A_S) = 2^{n-1}$*.*

5. *The star of* $L_{n,k}(a,b,-,-,d_1,d_2,e)$ *has complexity* $2^{n-2} + 2^{n-2-k} + 1$*. The maximal value* $2^{n-2} + 2^{n-3} + 1$ *is reached only when* $k = 1$*.*

6. $L'_{m,j}(a,b,c_1,-,d_1,d_2,e)L_{n,k}(a,d_2,c_1,-,d_1,b,e)$ *has complexity* $m-1-j+j2^{n-2}+2^{n-1}$*. The maximal value* $m2^{n-2}+1$ *is reached only when* $j = m-2$*.*

7. *For* $m,n \geqslant 3$*,* $1 \leqslant j \leqslant m-2$*, and* $1 \leqslant k \leqslant n-2$*, define the languages* $L'_{m,j} = L'_{m,j}(a,b,c_1,-,d_1,d_2,e)$ *and* $L_{n,k} = L_{n,k}(a,b,e,-,d_2,d_1,c_1)$*. For any proper binary boolean function* $\circ$*, the complexity of* $L'_{m,j} \circ L_{n,k}$ *is maximal. In particular,*

   (a) $L'_{m,j} \cup L_{n,k}$ *and* $L'_{m,j} \oplus L_{n,k}$ *have complexity* mn*.*

   (b) $L'_{m,j} \setminus L_{n,k}$ *has complexity* $mn-(n-1)$*.*

   (c) $L'_{m,j} \cap L_{n,k}$ *has complexity* $mn-(m+n-2)$*.*

Proof. The remainder of this paper is the proof of this theorem. The longer parts of the proof are separated into individual propositions and lemmas.

DFA $\mathcal{D}_{n,k}(a,b,-,-,-,d_2,e)$ is easily seen to be minimal. Language $L_{n,k}(\Sigma)$ is k-proper by Propositions 2 and 3.

1. See Lemma 5 and Proposition 6.

2. If the initial state of $\mathcal{D}_{n,k}(a,b,-,-,-,d_2,e)$ is changed to $q \in E_{n,k}$, the new DFA accepts a quotient of $L_{n,k}$ and is still minimal; hence the complexity of that quotient is n. If the initial state is changed to $q \in F_{n,k}$ then states in $E_{n,k}$ are unreachable, but the DFA on $\{n-1-k,\dots,n-1\}$ is minimal; hence the complexity of that quotient is $k+1$. The remaining quotient is empty, and hence has complexity 1. By Proposition 2, these are maximal.

3. See Proposition 7 for the reverse. It was shown in [9] that the number of atoms is equal to the complexity of the reverse.
4. See Proposition 8.
5. See Proposition 9.
6. See Proposition 10.
7. By [3, Theorem 2], all boolean operations on regular languages have the upper bound mn, which gives the bound for (a). The bounds for (b) and (c) follow from [3, Theorem 5]. Proposition 11 proves that all these bounds are tight for $L'_{m,j} \circ L_{n,k}$. □

Lemma 5. *Let $n \geqslant 1$ and $1 \leqslant k \leqslant n-2$. For any permutation t of Q_n such that $E_{n,k}t = E_{n,k}$, $F_{n,k}t = F_{n,k}$, and $(n-1)t = n-1$, there is a word $w \in \{a,b\}^*$ that induces t on $\mathcal{D}_{n,k}$.*

Proof. Only a and b induce permutations of Q_n; every other letter induces a properly injective map. Furthermore, a and b permute $E_{n,k}$ and $F_{n,k}$ separately, and both fix $n-1$. Hence every $w \in \{a,b\}^*$ induces a permutation on Q_n such that $E_{n,k}w = E_{n,k}$, $F_{n,k}w = F_{n,k}$, and $(n-1)w = n-1$. Each such permutation naturally corresponds to an element of $S_{n-1-k} \times S_k$, where S_m denotes the symmetric group on m elements. To be consistent with the DFA, assume S_{n-1-k} contains permutations of $\{0,\dots,n-2-k\}$ and S_k contains permutations of $\{n-1-k,\dots,n-2\}$. Let s_a and s_b denote the group elements corresponding to the transformations induced by a and b respectively. We show that s_a and s_b generate $S_{n-1-k} \times S_k$.

It is well known that $(0,\dots,m-1)$, and $(0,1)$ generate the symmetric group on $\{0,\dots,m-1\}$ for any $m \geq 2$. Note that $(1,\dots,m-1)$ and $(0,1)$ are also generators, since $(0,1)(1,\dots,m-1) = (0,\dots,m-1)$.

If $n-1-k = 1$ and $k = 1$, then $S_{n-1-k} \times S_k$ is the trivial group. If $n-1-k = 1$ and $k \geqslant 2$, then $s_a = (\mathbb{1}, (n-1-k, n-k))$ and s_b is either $(\mathbb{1}, (n-1-k,\dots,n-2))$ or $(\mathbb{1}, (n-k,\dots,n-2))$, and either pair generates the group. There is a similar argument when $k = 1$.

Assume now $n-1-k \geqslant 2$ and $k \geqslant 2$. If $n-1-k$ is odd then $s_a = ((0,\dots,n-2-k), (n-1-k, n-k))$, and hence $s_a^{n-1-k} = ((0,\dots,n-2-k)^{n-1-k}, (n-1-k,n-k)^{n-1-k}) = (\mathbb{1}, (n-1-k, n-k))$. Similarly if $n-1-k$ is even then $s_a = ((1,\dots,n-2-k), (n-1-k, n-k))$, and hence $s_a^{n-2-k} = (\mathbb{1}, (n-1-k, n-k))$. Therefore $(\mathbb{1}, (n-1-k, n-k))$ is always generated by s_a. By symmetry, $((0,1), \mathbb{1})$ is always generated by s_b regardless of the parity of k.

Since we can isolate the transposition component of s_a, we can isolate the other component as well: $(\mathbb{1}, (n-1-k, n-k))s_a$ is either $((0,\dots,n-2-k), \mathbb{1})$ or $((1,\dots,n-2-k), \mathbb{1})$. Paired with $((0,1), \mathbb{1})$, either element is sufficient to generate $S_{n-1-k} \times \{\mathbb{1}\}$. Similarly, s_a and s_b generate $\{\mathbb{1}\} \times S_k$. Therefore s_a and s_b generate $S_{n-1-k} \times S_k$. It follows that a and b generate all permutations t of Q_n such that $E_{n,k}t = E_{n,k}$, $F_{n,k}t = F_{n,k}$, and $(n-1)t = n-1$. □

Proposition 6 (Syntactic Semigroup). *The syntactic semigroup of $L_{n,k}(\Sigma)$ has cardinality $n^{n-1-k}(k+1)^k$, which is maximal for a k-proper language. Furthermore, seven letters are required to meet this bound. The maximum value $n(n-1)^{n-2}$ is reached only when $k = n-2$.*

Proof. Let L be a k-proper language of complexity n and let $\mathcal{D}$ be a minimal DFA recognizing L. By Lemma 1, $\mathcal{D}$ has an empty state. By Proposition 2, the only states that can be reached from one of the k final states are either final or empty. Thus, a transformation in the transition semigroup of $\mathcal{D}$ may map each final state to one of $k+1$ possible states, while each non-final, non-empty state may be mapped to any of the n states. Since the empty state can only be mapped to itself, we are left with $n^{n-1-k}(k+1)^k$ possible transformations in the transition semigroup. Therefore the syntactic semigroup of any k-proper language has size at most $n^{n-1-k}(k+1)^k$.

Now consider the transition semigroup of $\mathcal{D}_{n,k}(\Sigma)$. Every transformation t in the semigroup must satisfy $F_{n,k}t \subseteq F_{n,k} \cup \{n-1\}$ and $(n-1)t = n-1$, since any other transformation would violate prefix-convexity. We show that the semigroup contains every such transformation, and hence the syntactic semigroup of $L_{n,k}(\Sigma)$ is maximal.

First, consider the transformations t such that $E_{n,k}t \subseteq E_{n,k} \cup \{n-1\}$ and $qt = q$ for all $q \in F_{n,k} \cup \{n-1\}$. By Lemma 5, a and b generate every permutation of $E_{n,k}$. When t is not a permutation, we can use c_1 to combine any states p and q: apply a permutation on $E_{n,k}$ so that $p \to 0$ and $q \to 1$, and then apply c_1 so that $1 \to 0$. Repeat this method to combine any set of states, and further apply permutations to induce the desired transformation while leaving the states of $F_{n,k} \cup \{n-1\}$ in place. The same idea applies with d_1; apply permutations and d_1 to send any states of $E_{n,k}$ to $n-1$. Hence a, b, c_1, and d_1 generate every transformation t such that $E_{n,k}t \subseteq E_{n,k} \cup \{n-1\}$ and $qt = q$ for all $q \in F_{n,k} \cup \{n-1\}$.

We can make the same argument for transformations that act only on $F_{n,k}$ and fix every other state. Since c_2 and d_2 act on $F_{n,k}$ exactly as c_1 and d_1 act on $E_{n,k}$, the letters a, b, c_2, and d_2 generate every transformation t such that $F_{n,k}t \subseteq F_{n,k} \cup \{n-1\}$ and $qt = q$ for all $q \in E_{n,k} \cup \{n-1\}$. It follows that a, b, c_1, c_2, d_1, and d_2 generate every transformation t such that $E_{n,k}t \subseteq E_{n,k} \cup \{n-1\}$, $F_{n,k}t \subseteq F_{n,k} \cup \{n-1\}$, and $(n-1)t = n-1$.

Note the similarity between this DFA restricted to the states $E_{n,k} \cup \{n-1\}$ (or $F_{n,k} \cup \{n-1\}$) and the witness for right ideals introduced in [7]. The argument for the size of the syntactic semigroup of right ideals is similar to this; see [10].

Finally, consider an arbitrary transformation t such that $F_{n,k}t \subseteq F_{n,k} \cup \{n-1\}$ and $(n-1)t = n-1$. Let j_t be the number of states $p \in E_{n,k}$ such that $pt \in F_{n,k}$. We show by induction on j_t that t is in the transition semigroup of $\mathcal{D}$. If $j_t = 0$, then t is generated by $\Sigma \setminus \{e\}$. If $j_t \geqslant 1$, there exist $p, q \in E_{n,k}$ such that $pt \in F_{n,k}$ and q is not in the image of t. Consider the transformations s_1 and s_2 defined by $qs_1 = pt$ and $rs_1 = r$ for $r \neq q$, and $ps_2 = q$ and $rs_2 = rt$ for $r \neq p$. Then $(rs_2)s_1 = rt$ for all $r \in Q_n$. Notice that

$j_{s_2} = j_t - 1$, and hence Σ generates s_2 by inductive assumption. One can verify that $s_1 = (n-1-k, pt)(0,q)(0 \to n-1-k)(0,q)(n-1-k, pt)$. From this expression, we see that s_1 is the composition of transpositions induced by words in $\{a,b\}^*$ and the transformation $(0 \to n-1-k)$ induced by e, and hence s_1 is generated by Σ. Thus, t is in the transition semigroup. By induction on j_t, it follows that the syntactic semigroup of $L_{n,k}$ is maximal.

Now we show that seven letters are required to meet this bound. Two letters (like a and b) are required to generate the permutations, since clearly one letter is not sufficient. Every other letter will induce a properly injective map. A letter (like c_1) that induces a properly injective map on $E_{n,k}$ and permutes $F_{n,k}$ is required. Similarly, a letter (like c_2) that permutes $E_{n,k}$ and induces a properly injective map on $F_{n,k}$ is required. A letter (like d_1) that sends a state in $E_{n,k}$ to $n-1$ and permutes $F_{n,k}$ is required. Similarly, a letter (like d_2) that sends a state in $F_{n,k}$ to $n-1$ and permutes $E_{n,k}$ is required. Finally, a letter (like e) that connects $E_{n,k}$ and $F_{n,k}$ is required.

For a fixed n, we may want to know which $k \in \{1,\dots,n-2\}$ maximizes $s_n(k) = n^{n-1-k}(k+1)^k$; this corresponds to the largest syntactic semigroup of a proper prefix-convex language with n quotients. We show that $s_n(k)$ is largest at $k = n-2$. Consider the ratio $\frac{s_n(k+1)}{s_n(k)} = \frac{(k+2)^{k+1}}{n(k+1)^k}$. Notice this ratio is increasing with k, and hence s_n is a convex function on $\{1,\dots,n-2\}$. It follows that the maximum value of s_n must occur at one the endpoints, 1 and $n-2$.

Now we show that $s_n(n-2) \geqslant s_n(1)$ for all $n \geqslant 3$. We can check this explicitly for $n = 3,4,5$. When $n \geqslant 6$, $s_n(n-2)/s_n(1) = \frac{n}{2}\left(\frac{n-1}{n}\right)^{n-2} \geqslant 3\,(1/e) > 1$; so the largest syntactic semigroup of $L_{n,k}(\Sigma)$ occurs only at $k = n-2$ for all $n \geqslant 3$. □

Proposition 7 (Reverse). *For any regular language of complexity n with an empty quotient, the reversal has complexity at most 2^{n-1}. Moreover, the reverse of $L_{n,k}(a,b,-,-,-,d_2,e)$ has complexity 2^{n-1} for $n \geqslant 3$ and $1 \leqslant k \leqslant n-2$.*

Proof. The first claim is left for the reader to verify. For the second claim, let $\mathcal{D}_{n,k} = (Q_n, \{a,b,d_2,e\}, \delta_{n,k}, 0, F_{n,k})$ denote the DFA $\mathcal{D}_{n,k}(a,b,-,-,-,d_2,e)$ in Definition 2 and let $L_{n,k} = L(D_{n,k})$. Construct an NFA $\mathcal{N}$ recognizing the reverse of $L_{n,k}$ by reversing each transition, letting the initial state 0 be the unique final state, and letting the final states in $F_{n,k}$ be the initial states. Applying the subset construction to $\mathcal{N}$ yields a DFA $\mathcal{D}^R$ whose states are subsets of Q_{n-1}, with initial state $F_{n,k}$ and final states $\{G \subseteq Q_{n-1} \mid 0 \in G\}$. We show that $\mathcal{D}^R$ is minimal, and hence the reverse of $L_{n,k}$ has complexity 2^{n-1}.

Recall from Lemma 5 that a and b generate all permutations of $E_{n,k}$ and $F_{n,k}$ in $\mathcal{D}_{n,k}$. Although the transitions are reversed in $\mathcal{D}^R$, they still generate all such permutations. Let $u_1, u_2 \in \{a,b\}^*$ be such that u_1 induces $(0,\dots,n-2-k)$ and u_2 induces $(n-1-k,\dots,n-2)$ in $\mathcal{D}^R$.

Consider a state $U = \{q_1,\dots,q_h, n-1-k,\dots,n-2\}$ where $0 \leqslant q_1 < q_2 < \cdots < q_h \leqslant n-2-k$. If $h = 0$, then U is the initial state. When $h \geqslant 1$, $\{q_2-q_1, q_3-q_1,\dots,q_h-q_1, n-1-k,\dots,n-2\}eu_1^{q_1} = U$. By induction, all such states are reachable.

Now we show that any state $U = \{q_1, \dots, q_h, p_1, \dots, p_i\}$ where $0 \leqslant q_1 < q_2 < \cdots < q_h \leqslant n-2-k$ and $n-1-k \leqslant p_1 < p_2 < \cdots < p_i \leqslant n-2$ is reachable. If $i = k$, then $U = \{q_1, \dots, q_h, n-1-k, \dots, n-2\}$ is reachable by the argument above. When $0 \leqslant i < k$, choose $p \in F_{n,k} \setminus U$ and see that U is reached from $U \cup \{p\}$ by $u_2^{n-1-p} d_2 u_2^{p-(n-2-k)}$. By induction, every state is reachable.

To prove distinguishability, consider distinct states U and V. Choose $q \in U \oplus V$. If $q \in E_{n,k}$, then U and V are distinguished by $u_1^{n-1-k-q}$. When $q \in F_{n,k}$, they are distinguished by $u_2^{n-1-q}e$. So $\mathcal{D}^R$ is minimal. □

Proposition 8 (Atomic Complexity). *For each atom A_S of $L_{n,k}(\Sigma)$, write $S = X_1 \cup X_2$, where $X_1 \subseteq E_{n,k}$ and $X_2 \subseteq F_{n,k}$. Let $\overline{X_1} = E_{n,k} \setminus X_1$ and $\overline{X_2} = F_{n,k} \setminus X_2$. If $X_2 \neq \emptyset$, then $\kappa(A_S) =$*

$$1+\sum_{x_1=0}^{|X_1|}\sum_{x_2=1}^{|X_1|+|X_2|-x_1}\sum_{y_1=0}^{|\overline{X_1}|}\sum_{y_2=0}^{|\overline{X_1}|+|\overline{X_2}|-y_1}\binom{n-1-k}{x_1}\binom{k}{x_2}\binom{n-1-k-x_1}{y_1}\binom{k-x_2}{y_2}.$$

If $X_1 \neq \emptyset$ and $X_2 = \emptyset$, then $\kappa(A_S) =$

$$1+\sum_{x_1=0}^{|X_1|}\sum_{x_2=0}^{|X_1|-x_1}\sum_{y_1=0}^{|\overline{X_1}|}\sum_{y_2=0}^{k}\binom{n-1-k}{x_1}\binom{k}{x_2}\binom{n-1-k-x_1}{y_1}\binom{k-x_2}{y_2} - 2^k\sum_{y=0}^{|\overline{X_1}|}\binom{n-1-k}{y}.$$

Otherwise, $S = \emptyset$ and $\kappa(A_S) = 2^{n-1}$. The atomic complexity is maximal for k-proper languages.

Proof. Let L be a k-proper language with quotients $K_0, K_1, \dots, K_{n-1}$ where $K_0, \dots, K_{n-2-k}$ are non-final quotients, $K_{n-1-k}, \dots, K_{n-2}$ are final quotients, and $K_{n-1} = \emptyset$. For $S \subseteq Q_{n-1}$, we have $A_S = \bigcap_{i \in S} K_i \cap \bigcap_{i \in \overline{S}} \overline{K_i}$; note $n-1 \notin S$ since A_S must be non-empty.

The quotients are $w^{-1}A_S = \bigcap_{i \in S} w^{-1}K_i \cap \bigcap_{i \in \overline{S}} \overline{w^{-1}K_i}$. However $w^{-1}K_i$ is always another quotient K_j. Thus $w^{-1}A_S$ has the form $J_{T,U} = \bigcap_{i \in T} K_i \cap \bigcap_{i \in U} \overline{K_i}$ where $T = \{i \mid K_i = w^{-1}K_j \text{ for some } j \in S\}$ and $U = \{i \mid K_i = w^{-1}K_j \text{ for some } j \in \overline{S}\}$. For brevity, we write $S \xrightarrow{w} T$ and $\overline{S} \xrightarrow{w} U$; this notation is in agreement with the action of w on the states of $\mathcal{D}_{n,k}$ corresponding to S and $\overline{S}$.

Notice $n-1 \in U$ and if $T \cap U \neq \emptyset$ then $J_{T,U}$ is the empty quotient. Furthermore, for any word w, $J_{T,U} \xrightarrow{w} J_{Tw,Uw}$. To establish the upper bound, we just count the number of possible distinct $J_{T,U}$ for each value of S.

Write $S = X_1 \cup X_2$ where $X_1 \subseteq E_{n,k}$ and $X_2 \subseteq F_{n,k}$, and let $\overline{X_1} = E_{n,k} \setminus X_1$ and $\overline{X_2} = F_{n,k} \setminus X_2$. By Proposition 2 any word w maps X_1 to a subset of Q_n and X_2 to a subset of $F_{n,k} \cup \{n-1\}$. Similarly, w maps $\overline{X_1}$ to a subset of Q_n, $\overline{X_2}$ to a subset of $F_{n,k} \cup \{n-1\}$, and $n-1$ to itself.

One can bound the number of non-empty quotients of A_S by counting the number of disjoint $T, U \subseteq Q_n$ that could be reached from S and $\overline{S}$ respectively by some transformation in the transition semigroup. Specifically, we require $n-1 \in U$, $|T| \leqslant |S|$, $|U| \leqslant |\overline{S}|$, $|T \cap E_{n,k}| \leqslant |X_1|$, and $|U \cap E_{n,k}| \leqslant |\overline{X_1}|$. Thus we have the initial estimate

$$\sum_{x_1=0}^{|X_1|} \sum_{x_2=0}^{|X_1|+|X_2|-x_1} \sum_{y_1=0}^{|\overline{X_1}|} \sum_{y_2=0}^{|\overline{X_1}|+|\overline{X_2}|-y_1} \binom{n-1-k}{x_1}\binom{k}{x_2}\binom{n-1-k-x_1}{y_1}\binom{k-x_2}{y_2},$$

where x_1 counts $|T \cap E_{n,k}|$, x_2 counts $|T \cap F_{n,k}|$, y_1 counts $|U \cap E_{n,k}|$, and y_2 counts $|U \cap F_{n,k}|$. With some refinements, this estimate leads to the three cases in the statement.

Note if $S \neq \emptyset$ then $T \neq \emptyset$. Also, if $X_2 \neq \emptyset$, then any non-empty quotient $J_{T,U}$ must have $T \cap F_{n,k} \neq \emptyset$ since X_2 cannot be mapped to $n-1$. In the corresponding equation of the statement, this has the effect that x_2 cannot be 0. We must add 1 to account for the empty state, achieved when T and U intersect.

If $X_1 \neq \emptyset$ and $X_2 = \emptyset$, then we cannot have $x_1 = x_2 = 0$ since that would correspond to $T = \emptyset$; the subtracted term in the statement is the value of the estimate when $x_1 = x_2 = 0$. As before, add 1 for the empty quotient.

Finally, if $S = \emptyset$, then $T = \emptyset$ and $U \subseteq Q_n$ with $n-1 \in U$. There are 2^{n-1} possible values of U. Hence $\kappa(A_S) \leqslant 2^{n-1}$. There is no need to add 1 because T and U cannot intersect; there is not necessarily an empty quotient. This yields the three cases in the statement.

It remains to prove that $L_{n,k}(\Sigma)$ of Definition 2 meets this upper bound. Let the quotient K_q of $L_{n,k}$ be the language accepted by state q in $\mathcal{D}_{n,k}$. We must show that every $J_{T,U}$ can be reached from A_S by some word in Σ^*, and that every non-empty $J_{T,U}$ is distinct from $J_{T',U'}$ whenever $(T,U) \neq (T',U')$. By Proposition 6, the syntactic semigroup is as large as possible for k-proper languages. Hence, whenever $n-1 \in U$, $|T| \leqslant |S|$, $|U| \leqslant |\overline{S}|$, $|T \cap E_{n,k}| \leqslant |X_1|$, and $|U \cap E_{n,k}| \leqslant |\overline{X_1}|$, there is a word $w \in \Sigma^*$ such that $S \xrightarrow{w} T$ and $\overline{S} \xrightarrow{w} U$. Thus each quotient $J_{T,U}$ counted by the upper bound is reachable in A_S.

Consider $J_{T,U}$ where $T \cap U = \emptyset$ and $n-1 \in U$. If $T \neq \emptyset$ then there exists w such that $T \xrightarrow{w} \{n-2\}$ and $U \xrightarrow{w} \{n-1\}$; hence $w \in J_{T,U}$ since $\varepsilon \in K_{n-2}$. If $T = \emptyset$ choose w such that $U \xrightarrow{w} \{n-1\}$; hence $w \in J_{T,U}$. Thus $J_{T,U}$ is non-empty.

Now take $J_{T',U'}$ where $(T,U) \neq (T',U')$, $T' \cap U' = \emptyset$ and $n-1 \in U'$. We must show that $J_{T,U}$ and $J_{T',U'}$ are distinct. If $r \in T' \setminus T$, then choose w that maps $r \to n-1$ in $\mathcal{D}_{n,k}$; $J_{Tw,Uw}$ is non-empty, since $Tw \cap Uw = \emptyset$, and $J_{T'w,U'w} = \emptyset$ since $n-1 \in T'w$. Similarly, if $T = T'$ and $r \in U' \setminus U$, then choose w that maps $T \cup \{r\} \to \{n-2\}$ and $Q_n \setminus (T \cup \{r\}) \to \{n-1\}$. Then $J_{Tw,Uw} = J_{\{n-2\},\{n-1\}}$ is non-empty and $J_{T'w,U'w} = J_{\{n-2\},\{n-2,n-1\}} = \emptyset$. Finally, if $T = T' = \emptyset$ and $r \in U' \setminus U$, then distinguish $J_{T,U}$ and $J_{T',U'}$ by a word that sends $r \to n-2$ and $Q_n \setminus \{r\} \to \{n-1\}$. Hence, $J_{T,U}$ and $J_{T',U'}$ are distinct. Therefore, the quotients of A_S counted in the upper bound are pairwise distinct and $L_{n,k}$ has maximal atomic complexity. $\square$

Proposition 9 (Star). *Let L be a regular language with $n \geqslant 2$ quotients, including $k \geqslant 1$ final quotients and one empty quotient. Then $\kappa(L^*) \leqslant 2^{n-2} + 2^{n-2-k} + 1$. This bound is tight for proper prefix-convex languages; in particular, the language $(L_{n,k}(a, b, -, -, d_1, d_2, e))^*$ meets this bound for $n \geqslant 3$ and $1 \leqslant k \leqslant n-2$.*

Proof. Since L has an empty quotient, let $n-1$ be the empty state of its minimal DFA $\mathcal{D}$. To obtain an ε-NFA for L^*, we add a new initial state $0'$ which is final and has the same transitions as 0. We then add an ε-transition from every state in F to 0. Applying the subset construction to this ε-NFA yields a DFA $\mathcal{D}' = (Q', \Sigma, \delta', \{0'\}, F')$ recognizing L^*, in which Q' contains non-empty subsets of $Q_n \cup \{0'\}$.

Many of the states of Q' are unreachable or indistinguishable from other states. Since there is no transition in the ε-NFA to $0'$, the only reachable state in Q' containing $0'$ is $\{0'\}$. As well, any reachable final state $U \neq \{0'\}$ must contain 0 because of the ε-transitions. Finally, for any $U \in Q'$, we have $U \in F'$ if and only if $U \cup \{n-1\} \in F'$, and since $\delta'(U \cup \{n-1\}, w) = \delta'(U, w) \cup \{n-1\}$ for all $w \in \Sigma^*$, the states U and $U \cup \{n-1\}$ are equivalent in D'.

Hence $\mathcal{D}'$ is equivalent to a DFA with the states $\{\{0'\}\} \cup \{U \subseteq Q_{n-1} \mid U \cap F = \emptyset\} \cup \{U \subseteq Q_{n-1} \mid 0 \in U \text{ and } U \cap F \neq \emptyset\}$. This DFA has $1 + 2^{n-1-k} + (2^{n-2} - 2^{n-2-k}) = 2^{n-2} + 2^{n-2-k} + 1$ states. Thus, $\kappa(L^*) \leqslant 2^{n-2} + 2^{n-2-k} + 1$.

This bound must apply when L is a prefix-convex language and $n \geqslant 3$: by Lemma 1, L is either a right ideal or has an empty state. If L is a right ideal, then $\kappa(L^*) \leqslant n+1$, which is at most $2^{n-2} + 2^{n-2-k} + 1$ for $n \geqslant 3$.

For the last claim, let $\mathcal{D}_{n,k}(a, b, -, -, d_1, d_2, e)$ of Definition 2 be denoted by $\mathcal{D}_{n,k} = (Q_n, \{a, b, d_1, d_2, e\}, \delta_{n,k}, 0, F_{n,k})$ and let $L_{n,k} = L(D_{n,k})$. We apply the same construction and reduction as before to obtain a DFA $\mathcal{D}'_{n,k}$ recognizing $L^*_{n,k}$ with states $Q' = \{\{0'\}\} \cup \{U \subseteq E_{n,k}\} \cup \{U \subseteq Q_{n-1} \mid 0 \in U \text{ and } U \cap F_{n,k} \neq \emptyset\}$. We show that the states of Q' are reachable and pairwise distinguishable.

By Lemma 5, a and b generate all permutations of $E_{n,k}$ and $F_{n,k}$ in $\mathcal{D}_{n,k}$. Choose $u_1, u_2 \in \{a, b\}^*$ such that u_1 induces $(0, \dots, n-2-k)$ and u_2 induces $(n-1-k, \dots, n-2)$ in $\mathcal{D}_{n,k}$.

For reachability, we consider three cases. (1) State $\{0'\}$ is reachable by ε. (2) Let $U \subseteq E_{n,k}$. For any $q \in E_{n,k}$, we can reach $U \setminus \{q\}$ by $u_1^{n-2-k-q} d_1 u_1^q$; hence if U is reachable, then every subset of U is reachable. Observe that state $E_{n,k}$ is reachable by $e u_1^{n-2-k} d_2^k$, and we can reach any subset of this state. Therefore, all non-final states are reachable. (3) If $U \cap F_{n,k} \neq \emptyset$, then $U = \{0, q_1, q_2, \dots, q_h, r_1, \dots, r_i\}$ where $0 < q_1 < \dots < q_h \leqslant n-2-k$ and $n-1-k \leqslant r_1 < \dots < r_i < n-1$ and $i \geqslant 1$. We prove that U is reachable by induction on i. If $i = 0$, then U is reachable by (2). For any $i \geqslant 1$, we can reach U from $\{0, q_1, \dots, q_h, r_2 - (r_1 - (n-1-k)), \dots, r_i - (r_1 - (n-1-k))\}$ by $e u_2^{r_1 - (n-1-k)}$. Therefore, all states of this form are reachable.

Now we show that the states are pairwise distinguishable. (1) The initial state $\{0'\}$ is distinguishable from any other final state U since $\{0'\}u_1$ is non-final and Uu_1 is final. (2) If U and V are distinct subsets of $E_{n,k}$, then there is some $q \in U \oplus V$. We distinguish U and V by $u_1^{n-1-k-q} e$. (3) If U and V are distinct and final and neither one is $\{0'\}$, then there is some $q \in U \oplus V$. If $q \in E_{n,k}$, then $Ud_2^k = U \setminus F_{n,k}$ and $Vd_2^k = V \setminus F_{n,k}$ are distinct, non-final states as in (2). Otherwise, $q \in F_{n,k}$ and we distinguish U and V by $u_2^{n-1-q} d_2^{k-1}$. □

Proposition 10 (Product). *For $m, n \geqslant 3$, $1 \leqslant j \leqslant m-2$, and $1 \leqslant k \leqslant n-2$, the product of $L'_{m,j}(a, b, c_1, -, d_1, d_2, e)$ and $L_{n,k}(a, d_2, c_1, -, d_1, b, e)$ has complexity $m - 1 - j + j2^{n-2} + 2^{n-1}$.*

PROOF. Let $\mathcal{D}'_{m,j}$ and $\mathcal{D}_{n,k}$ be the DFAs of Definition 2 for $L'_{m,j}(a, b, c_1, -, d_1, d_2, e)$ and $L_{n,k}(a, d_2, c_1, -, d_1, b, e)$ respectively. As before, take $\mathcal{D}'_{m,j}$ to have the states $Q'_m = \{0', 1', \dots, (m-1)'\}$ and let $E'_{n,k} = \{0', \dots, (m-2-j)'\}$. Using the standard construction of the ε-NFA $\mathcal{N}$ for the product, we delete the empty state $n-1$, change the final states of $\mathcal{D}'_{m,j}$ to non-final states, and add ε-transitions from each final state of $\mathcal{D}'_{m,j}$ to the initial state of $\mathcal{D}_{n,k}$.

The subset construction on $\mathcal{N}$ yields states of the form $\{p'\} \cup S$, where $p' \in Q'_m$ and $S \subseteq Q_{n-1}$. However, some of these sets are not reachable in the product: if $p' \in E'_{m,j}$ then we must have $S = \emptyset$, and if $p' \in F'_{m,j}$ then $0 \in S$ because of the ε-transitions in $\mathcal{N}$.

Thus, we have the states $\{p'\}$ for $p' \in E'_{m,j}$, $\{p', 0\} \cup S$ for $p' \in F'_{m,j}$ and $S \subseteq Q_{n-1} \setminus \{0\}$, and $\{(m-1)'\} \cup S$ for $S \subseteq Q_{n-1}$. This totals to $(m-1-j) + (j2^{n-2}) + (2^{n-1}) = m - 1 - j + j2^{n-2} + 2^{n-1}$ different states. We show that they are reachable and pairwise distinguishable.

State $\{p'\}$ is reached by d_1^p for all $p' \in E'_{m,j}$. State $\{(m-1-j)', 0\}$ is reached by e. For $m - j \leqslant p \leqslant m-1$ we have $\{(m-1-j)', 0\} \xrightarrow{d_2^{p-(m-1-j)}}$ $\begin{cases} \{p', 0, 1\} & \text{if } n-1-k \geqslant 2 \\ \{p', 0\} & \text{if } n-1-k = 1 \end{cases} \xrightarrow{c_1} \{p', 0\}$.

Now consider states of the form $\{p', 0\} \cup T$ where $p' \in F'_{m,j}$ and $T \subseteq F_{n,k}$. These states are reachable when $T = \emptyset$. Inductively assume the states are reachable when $|T| < i$ for some $i \geqslant 1$. Let $T_i = \{r_1, r_2, \dots, r_i\}$ where $n-1-k \leqslant r_1 < r_2 < \cdots < r_i \leqslant n-2$, and let $T_{i-1} = \{r_2 - (r_1 - (n-1-k)), \dots, r_i - (r_1 - (n-1-k))\}$. Then $\{0\} \cup T_{i-1} \xrightarrow{e} \{n-1-k\} \cup T_{i-1} \xrightarrow{b^{r_1-(n-1-k)}} T_i$. Notice b induces a permutation on $\mathcal{D}'_{m,j}$, so for any $p' \in F'_{m,j}$ there is a state $q' \in F'_{m,j}$ such that $q' \xrightarrow{eb^{r_1-(n-1-k)}} p'$. Thus, $\{p', 0\} \cup T_i$ is reachable from $\{q', 0\} \cup T_{i-1}$.

Extend this to states of the form $\{p', 0\} \cup S \cup T$, where $p' \in F'_{m,j}$, $S \subseteq E_{n,k} \setminus \{0\}$, and $T \subseteq F_{n,k}$. These states are reachable when $S = \emptyset$. Inductively assume the states are reachable when $|S| < h$ for some $h \geqslant 1$. Let $S_h = \{q_1, q_2, \dots, q_h\}$ where $1 \leqslant q_1 < q_2 < \cdots < q_i \leqslant n-2-k$, and let $S_{h-1} = \{q_2 - q_1, \dots, q_h - q_1\}$. Then $\{p', 0\} \cup S_{h-1} \cup T \xrightarrow{d_1} \{p', 0, 1\} \cup (S_{h-1} + 1) \cup T \xrightarrow{(d_1 c_1)^{q_1 - 1}} \{p', 0, q_1\} \cup (S_{h-1} + q_1) \cup T = \{p', 0\} \cup S_h \cup T$. In the last derivation, $S + c$ denotes the set $\{q + c \colon q \in S\}$.

State $\{(m-1)', 0\} \cup S \cup T$ is reachable from $\{(m-2)', 0\} \cup S \cup T$ by d_2^ℓ, where $\ell > 0$ is the order of d_2 in $\mathcal{D}_{n,k}$ (i.e. d_2^ℓ induces the identity transformation on $\mathcal{D}_{n,k}$).

Finally, state $\{(m-1)'\} \cup S \cup T$ is reachable from $\{(m-1)', 0\} \cup S \cup T$: by Lemma 5, the permutation $(0, 1, \ldots, n-2-k)$ of $\mathcal{D}_{n,k}$ is generated by some $u_1 \in \{a, d_2\}^*$, and $\{(m-1)', 0\} \cup S \cup T \xrightarrow{u_1^{n-2-k}} \{(m-1)', n-2-k\} \cup (S-1) \cup T \xrightarrow{d_1} \{(m-1)'\} \cup S \cup T$. Thus all states are reachable.

We now check distinguishability in cases. Using Lemma 5, take words $u_1, u_2 \in \{a, d_2\}^*$ such that u_1 induces $(0, 1, \ldots, n-2-k)$ and u_2 induces $(n-1-k, n-k, \ldots, n-2)$ on $\mathcal{D}_{n,k}$. Note u_1 and u_2 act on $\mathcal{D}'_{m,j}$ as well.

1. Let $U = \{(m-1)'\}$ and let V be any other state. Notice U is the empty state. We show that V is non-empty.
   (a) If $q \in V \cap Q_{n-1}$ then by the minimality of $\mathcal{D}_{n,k}$ there is a word w such that $qw \in F_{n,k}$; hence Vw is final.
   (b) Otherwise $V = \{p'\}$ for some $p' \in E'_{m,j}$. There is a word w such that $p'w \in F'_{m,j}$; hence $0 \in Vw$ and this reduces to Case (a).
2. Let $U = \{p'\}$ and $V = \{q'\}$ where $p', q' \in E'_{m,j}$ and $p < q$. Then $Vd_1^{m-1-j-q} = \{(m-1)'\}$ and $Ud_1^{m-1-j-q}$ is non-empty by Case 1.
3. Let $U = \{p'\}$ and $V = \{q', 0\} \cup S$ where $p' \in E'_{m,j}$, $q' \in F'_{m,j}$, and $S \subseteq Q_{n-1} \setminus \{0\}$. Then U and V are distinguished by e.
4. Let $U = \{p'\}$ and $V = \{(m-1)'\} \cup S$ where $p' \in E'_{m,j}$ and $S \subseteq Q_{n-1}$. If $S = \emptyset$ this reduces to Case 1. If $S \cap F_{n,k} \neq \emptyset$ then V is final. Otherwise there is some $r \in S$, and $Vu_1^{n-1-k-r}e$ is final. Notice $Uu_1^{n-1-k-r}e$ is non-final because $u_1 \in \{a, d_2\}^*$.
5. Let $U = \{(m-1)'\} \cup S$ and $V = \{(m-1)'\} \cup T$ where $S \neq T \subseteq Q_{n-1}$; pick $r \in S \oplus T$. Without loss of generality, say $r \in S \setminus T$.
   (a) If $r = 0$, then $U \xrightarrow{b^k} U \setminus F_{n,k} \xrightarrow{e} U \setminus (\{0\} \cup F_{n,k}) \cup \{n-1-k\}$ and $V \xrightarrow{b^k} V \setminus F_{n,k} \xrightarrow{e} V \setminus F_{n,k}$.
   (b) If $r \in E_{n,k}$, then we reduce to Case (a) by applying $u_1^{n-1-k-r}$.
   (c) If $r = n-1-k$, then Ub^{k-1} is final and Vb^{k-1} is non-final.
   (d) If $r \in F_{n,k}$, then we reduce to Case (c) by applying u_2^{n-1-r}.
6. Let $U = \{p', 0\} \cup S$ and $V = \{(m-1)'\} \cup T$ where $p' \in F'_{m,j}$, and $S, T \subseteq Q_{n-1}$. Notice $Ud_1^{n-1-k}b^k$ is non-empty since p' is not mapped to $(m-1)'$, but $V \xrightarrow{d_1^{n-1-k}} \{(m-1)'\} \cup T \setminus E_{n,k} \xrightarrow{b^k} \{(m-1)'\}$; this reduces to Case 1.
7. Let $U = \{p', 0\} \cup S$ and $V = \{q', 0\} \cup T$ where $p', q' \in F'_{m,j}$, $p < q$, and $S, T \subseteq Q_{n-1}$. Reduce to Case 6 by applying d_2^{m-1-q}.
8. Let $U = \{p', 0\} \cup S$ and $V = \{p', 0\} \cup T$ where $p' \in F'_{m,j}$ and $S \neq T \subseteq Q_{n-1}$. Pick $r \in S \oplus T$ and assume without loss of generality that $r \in S \setminus T$.

(a) If $r \geqslant 2$, then d_2^{m-1-p} fixes r and maps p' to $(m-1)'$; hence this reduces to Case 5.

(b) If $p = m-2$, then apply d_2 to reduce to Case 5. Notice Sd_2 and Td_2 are distinct since d_2 induces a permutation on $\mathcal{D}_{n,k}$.

(c) If $r = 1$ and $n-1-k \geqslant 2$, then applying d_1 reduces to Case (a).

(d) If $r = 1$ and $n-1-k = 2$, then observe that a and b both fix 1 in $\mathcal{D}_{n,k}$. By Lemma 5, there is a word $w \in \{a,b\}^*$ such that $p'w = (m-2)'$. Since $n-1-k = 2$, a and b do not alter $E_{n,k}$. Hence $1 \in Sw$ and $1 \notin Tw$, so this reduces to Case (b). □

Proposition 11 (Boolean Operations). *For $m, n \geqslant 3$, $1 \leqslant j \leqslant m-2$, and $1 \leqslant k \leqslant n-2$, let $L'_{m,j} = L'_{m,j}(a,b,c_1,-,d_1,d_2,e)$ and let $L_{n,k} = L_{n,k}(a,b,e,-,d_2,d_1,c_1)$ of Definition 2. For any proper binary boolean function $\circ$, the complexity of $L'_{m,j} \circ L_{n,k}$ is maximal. In particular,*

1. *$\kappa(L'_{m,j} \cup L_{n,k}) = \kappa(L'_{m,j} \oplus L_{n,k}) = mn$.*
2. *$\kappa(L'_{m,j} \setminus L_{n,k}) = mn-(n-1)$.*
3. *$\kappa(L'_{m,j} \cap L_{n,k}) = mn-(m+n-2)$.*

Proof. Let $\mathcal{D}'_{m,j}$ and $\mathcal{D}_{n,k}$ be the DFAs of Definition 2 for $L'_{m,j}(a,b,c_1,-,d_1,d_2,e)$ and $L_{n,k}(a,b,e,-,d_2,d_1,c_1)$ respectively. As before, take $\mathcal{D}'_{m,j}$ to have the states $Q'_m = \{0',1',\dots,(m-1)'\}$. There is a standard construction for $L'_{m,j} \circ L_{n,k}$ for any boolean set operation $\circ$ in terms of the direct product. The direct product of $\mathcal{D}'_{m,j}$ and $\mathcal{D}_{n,k}$ has states $Q'_m \times Q_n$, initial state $(0',0)$, and transition function δ such that $\delta((p',q),w) = (\delta'_{m,j}(p',w), \delta_{n,k}(q,w))$. If we set the final states to be $(F'_{m,j} \times Q_n) \circ (Q'_m \times F_{n,k})$, it is a DFA recognizing $L'_{m,j} \circ L_{n,k}$. For each $\circ \in \{\cup, \oplus, \setminus, \cap\}$, we construct the DFA $\mathcal{D}_\circ$ to recognize $L'_{m,j} \circ L_{n,k}$. All four DFAs have the same states and transitions as the direct product and will only differ in the set of final states. The DFA $\mathcal{D}_\oplus$ for symmetric difference is shown in Figure 4.

We can usefully partition the states of the direct product. Let $W = E'_{m,j} \times E_{n,k}$, $X = E'_{m,j} \times F_{n,k}$, $Y = F'_{m,j} \times E_{n,k}$, $Z = F'_{m,j} \times F_{n,k}$, and $S = W \cup X \cup Y \cup Z$. Let $R = \{(m-1)'\} \times Q_n$ and $C = Q'_m \times \{n-1\}$.

We check that every state in the direct product is reachable. Since $\mathcal{D}_\cup$, $\mathcal{D}_\oplus$, $\mathcal{D}_\setminus$, and $\mathcal{D}_\cap$ have the same structure as the direct product, this argument will apply to them as well. By Lemma 5 there exist $u_1, u_2 \in \{a,b\}^*$ such that u_1 induces $(0',\dots,(m-2-j)')$ and u_2 induces $((m-1-j)',\dots,(m-1)')$ in $\mathcal{D}'_{m,j}$. Note that u_1 and u_2 permute $E_{n,k}$ and $F_{n,k}$ in $\mathcal{D}_{n,k}$. Similarly, there exist $v_1, v_2 \in \{a,b\}^*$ such that v_1 induces $(0,\dots,n-2-k)$ and v_2 induces $(n-1-k,\dots,n-1)$ in $\mathcal{D}_{n,k}$, and they permute $E'_{m,j}$ and $F'_{m,j}$ in $\mathcal{D}'_{m,j}$.

1. State $(p',q) \in W$ is reachable since $(0',0) \xrightarrow{d_1^p} (p',0) \xrightarrow{d_2^q} (p',q)$.

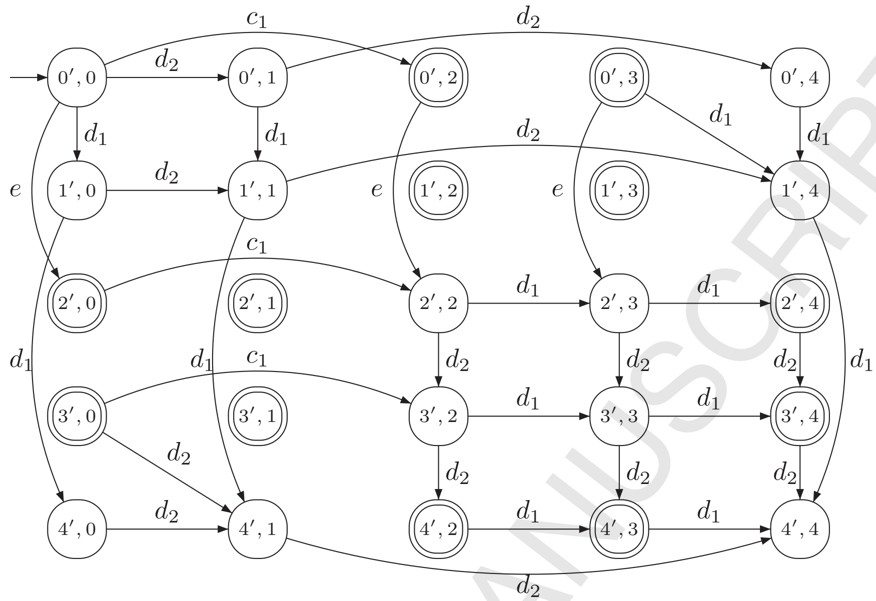

Figure 4: DFA $\mathcal{D}_{\oplus}$ for symmetric difference of proper languages with DFAs $\mathcal{D}'_{5,2}(a, b, c_1, -, d_1, d_2, e)$ and $\mathcal{D}_{5,2}(a, b, e, -, d_2, d_1, c_1)$ shown partially.

2. State $(p', 0) \in Y$ is reachable since $(0', 0) \xrightarrow{e} ((m-1-j)', 0) \xrightarrow{(d_2 e)^{p-(m-1-j)}} (p', 0)$. An arbitrary $(p', q) \in Y$ is then reached by v_1^q from some $(r', 0)$ where $r' \in F'_{m,j}$ is chosen so that $r' \xrightarrow{v_1^q} p'$ in $\mathcal{D}'_{m,j}$.
3. State $(p', q) \in X$ is reachable by symmetry with Case 2.
4. State $(p', q) \in Z$ is reachable since $(0', 0) \xrightarrow{ec_1} ((m-1-j)', n-1-k) \xrightarrow{d_2^{p-(m-1-j)}} (p', n-1-k) \xrightarrow{d_1^{q-(n-1-k)}} (p', q)$.
5. State $(p', n-1) \in C$ is reachable since $(0', 0) \xrightarrow{d_2^{n-1-k}} (0', n-1)$, and p' is reachable in $\mathcal{D}'_{m,j}$.
6. State $((m-1)', q) \in R$ is reachable by symmetry with Case 5.

Hence all states are reachable.

As a tool for distinguishability, we show that the states of S are distinguishable with respect to $R \cup C$; that is, for any pair of distinct states in S, we show that there is a word that sends one state to $R \cup C$ and leaves the other state in S. We check this fact in cases. Note that d_2 fixes the states of X and d_1 fixes the states of Y.

1. States of W and X are distinguished by words in d_2^*.
2. States of W and Y are distinguished by words in d_1^*.

3. States of X and Y are distinguished by words in d_1^*.

4. States of X and Z are distinguished by words in d_2^*.

5. States of Y and Z are distinguished by words in d_1^*.

6. To distinguish states of W and Z, we reduce to Case 5 by a word in $u_1^* e$.

7. Any two states of W are distinguished by a word in d_1^* if they differ in the first coordinate, or by a word in d_2^* if they differ in the second coordinate.

8. Any two states of Z are distinguished by a word in d_2^* if they differ in the first coordinate, or by a word in d_1^* if they differ in the second coordinate.

9. To distinguish two states of X, reduce to Case 4 by a word in $u_1^* e$ if they differ in the first coordinate, or reduce to Case 8 by a word in $u_1^* e$ if the first coordinate is the same.

10. Any two states of Y are distinguishable by symmetry with Case 9.

Now we determine which states are pairwise distinguishable with respect to the final states of $\mathcal{D}_\circ$ for each $\circ \in \{\cup, \oplus, \setminus, \cap\}$. Let $w = (u_1 e)^{m-1-j}(v_1 c_1)^{n-1-k}$; observe that w maps every state of S to a state of Z.

$\cup$, $\oplus$: In $\mathcal{D}_\cup$, (p', q) is final if $p' \in F'_{m,j}$ or $q \in F_{n,k}$. In $\mathcal{D}_\oplus$, (p', q) is final if $p' \in F'_{m,j}$ and $q \notin F_{n,k}$ or $p' \notin F'_{m,j}$ and $q \in F_{n,k}$. We show that all mn states are pairwise distinguishable in both cases.

The states of R are pairwise distinguishable by the minimality of $\mathcal{D}_{n,k}$. Similarly, the states of C are pairwise distinguishable by the minimality of $\mathcal{D}'_{m,j}$. The states of R and C are distinguishable by wd_1^k, since $R \setminus \{((m-1)', n-1)\} \xrightarrow{w} \{(m-1)'\} \times F_{n,k} \xrightarrow{d_1^k} \{(m-1)', n-1\}$ and $C \setminus \{((m-1)', n-1)\} \xrightarrow{w} F'_{m,j} \times \{n-1\} \xrightarrow{d_1^k} F'_{m,j} \times \{n-1\}$. The states of C and S are distinguishable since $S \xrightarrow{w} Z \xrightarrow{d_2^j} \{(m-1)'\} \times F_{n,k} \subseteq R$, and we can distinguish states of R and C. The states of R and S are similarly distinguishable. Finally, states of S are pairwise distinguishable because they can be distinguished with respect to $R \cup C$, and we can distinguish states of S and $R \cup C$.

$\setminus$: In $\mathcal{D}_\setminus$, (p', q) is final if $p' \in F'_{m,j}$ and $q \notin F_{n,k}$. The states of R are all empty, and the remaining states are pairwise distinguishable for a total of $mn - (n-1)$ distinguishable states.

The states of C are pairwise distinguishable by the minimality of $\mathcal{D}'_{m,j}$. The states of C and S are distinguishable since $S \xrightarrow{w} Z \xrightarrow{d_2^j} \{(m-1)'\} \times F_{n,k} \subseteq R$, and every state in R is empty. Finally, states of S are pairwise distinguishable because they can be distinguished with respect to $R \cup C$, and we can distinguish states of S and $R \cup C$.

$\cap$: In $\mathcal{D}_\cap$ the final state set is Z. The states of $R \cup C$ are all empty, leaving $mn - (m+n+2)$ distinguishable states. The states of S are non-empty since $S \xrightarrow{w} Z$. We can distinguish the states of S with respect to $R \cup C$; hence they are pairwise distinguishable. □

3. Conclusions

The bounds for prefix-convex languages (see also [8]) are summarized in Table 1. The largest bounds are shown in boldface type, and they are reached either in the class of right-ideal languages or the class of proper languages. Recall that for regular languages we have the following results: semigroup n^n, reverse 2^n, star $2^{n-1} + 2^{n-2}$, product $m2^n - 2^{n-1}$, boolean operations mn.

Table 1: Complexity of prefix-convex languages. For proper languages, the variables j and k refer to the number of final quotients of the languages of complexity m and n, respectively.

	Right-Ideal	Prefix-Closed	Prefix-Free	Proper
SeGr	$\mathbf{n^{n-1}}$	$\mathbf{n^{n-1}}$	n^{n-2}	$n^{n-1-k}(k+1)^k$
Rev	$\mathbf{2^{n-1}}$	$\mathbf{2^{n-1}}$	$2^{n-2}+1$	$\mathbf{2^{n-1}}$
Star	$n+1$	$2^{n-2}+1$	n	$\mathbf{2^{n-2}+2^{n-2-k}+1}$
Prod	$m+2^{n-2}$	$(m+1)2^{n-2}$	$m+n-2$	$\mathbf{m-1-j+j2^{n-2}+2^{n-1}}$
$\cup$	$mn-(m+n-2)$	$\mathbf{mn}$	$mn-2$	$\mathbf{mn}$
$\oplus$	$\mathbf{mn}$	$\mathbf{mn}$	$mn-2$	$\mathbf{mn}$
$\setminus$	$\mathbf{mn-(m-1)}$	$\mathbf{mn-(n-1)}$	$mn-(m+2n-4)$	$\mathbf{mn-(n-1)}$
$\cap$	$\mathbf{mn}$	$mn-(m+n-2)$	$mn-2(m+n-3)$	$mn-(m+n-2)$

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

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

[3] J.A. Brzozowski, Quotient complexity of regular languages, J. Autom. Lang. Comb. 15 (2010) 71–89.

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

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

[6] J.A. Brzozowski, G. Jirásková, C. Zou, Quotient complexity of closed languages, Theory Comput. Syst. 54 (2014) 277–292.

[7] J.A. Brzozowski, C. Sinnamon, Complexity of prefix-convex regular languages, 2016. http://arxiv.org/abs/1605.06697.

[8] J.A. Brzozowski, C. Sinnamon, Complexity of right-ideal, prefix-closed, and prefix-free regular languages., Acta Cybern. 23 (2017) 9–41.

[9] J.A. Brzozowski, H. Tamm, Theory of átomata, Theoret. Comput. Sci. 539 (2014) 13–27.

[10] J.A. Brzozowski, Y. Ye, Syntactic complexity of ideal and closed languages, in: G. Mauri, A. Leporati (Eds.), DLT 2011, volume 6795 of *LNCS*, Springer Berlin / Heidelberg, 2011, pp. 117–128.

[11] M. Crochemore, C. Hancart, Automata for pattern matching, in: G. Rozenberg, A. Salomaa (Eds.), Handbook of Formal Languages, volume 2, Springer, 1997, pp. 399–462.

[12] Y. Gao, N. Moreira, R. Reis, S. Yu, A survey on operational state complexity, J. Autom. Lang. Comb. 21 (2016) 251–310.

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

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

[15] B. Krawetz, J. Lawrence, J. Shallit, State complexity and the monoid of transformations of a finite set, in: M. Domaratzki, et al. (Eds.), CIAA 2005, volume 3317 of *LNCS*, Springer Berlin / Heidelberg, 2005, pp. 213–224.

[16] A.N. Maslov, Estimates of the number of states of finite automata, Dokl. Akad. Nauk SSSR 194 (1970) 1266–1268 (Russian). English translation: Soviet Math. Dokl. **11** (1970) 1373–1375.

[17] G. Thierrin, Convex languages, in: M. Nivat (Ed.), Automata, Languages and Programming, North-Holland, 1973, pp. 481–492.

[18] S. Yu, State complexity of regular languages, J. Autom. Lang. Comb. 6 (2001) 221–234.

[19] S. Yu, Q. Zhuang, K. Salomaa, The state complexities of some basic operations on regular languages, Theoret. Comput. Sci. 125 (1994) 315–328.