

Quotient Complexities of Atoms in Regular Ideal Languages*

Janusz Brzozowski[†] and Sylvie Davies[‡]

Abstract

A (left) quotient of a language L by a word w is the language $w^{-1}L = \{x \mid wx \in L\}$. The quotient complexity of a regular language L is the number of quotients of L ; it is equal to the state complexity of L , which is the number of states in a minimal deterministic finite automaton accepting L . An atom of L is an equivalence class of the relation in which two words are equivalent if for each quotient, they either are both in the quotient or both not in it; hence it is a non-empty intersection of complemented and uncomplemented quotients of L . A right (respectively, left and two-sided) ideal is a language L over an alphabet Σ that satisfies $L = L\Sigma^*$ (respectively, $L = \Sigma^*L$ and $L = \Sigma^*L\Sigma^*$). We compute the maximal number of atoms and the maximal quotient complexities of atoms of right, left and two-sided regular ideals.

Keywords: atom, left ideal, quotient, quotient complexity, regular language, right ideal, state complexity, syntactic semigroup, two-sided ideal

I dedicate this work to the memory of Ferenc Gécseg. I have known Ferenc for many years not only as an eminent scientist, author of numerous publications, and editor of Acta Cybernetica, but also as a good friend. I fondly remember his generosity and hospitality during my frequent visits to Hungary.

Janusz Brzozowski

1 Introduction

We assume that the reader is familiar with basic concepts of regular languages and finite automata; more background is given in the next section. Consider a regular language L over a finite non-empty alphabet Σ . Let $\mathcal{D} = (Q, \Sigma, \delta, q_1, F)$ be a minimal *deterministic finite automaton (DFA)* recognizing L , where Q is the set of states, $\delta: Q \times \Sigma \rightarrow Q$ is the *transition function*, q_1 is the *initial state*, and $F \subseteq Q$

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

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

[‡]Department of Pure Mathematics, University of Waterloo, Waterloo, ON, Canada N2L 3G1. E-mail: sldavies@uwaterloo.ca

is the set of *final* states. There are three natural equivalence relations associated with L and \mathcal{D} .

The *Nerode right congruence* [14] is defined as follows: Two words x and y are equivalent if for every $v \in \Sigma^*$, xv is in L if and only if yv is in L . The set of all words that “can follow” a given word x in L is the *left quotient of L by x* , defined by $x^{-1}L = \{v \mid xv \in L\}$. In automaton-theoretic terms $x^{-1}L$ is the set of all words v that are accepted from the state $q = \delta(q_1, x)$ reached when x is applied to the initial state of \mathcal{D} ; this is known as the *right language* of state q , the language accepted by DFA $\mathcal{D}_q = (Q, \Sigma, \delta, q, F)$. The Nerode equivalence class containing x is known as the *left language* of state q , the language accepted by DFA ${}_q\mathcal{D} = (Q, \Sigma, \delta, q_1, \{q\})$. The number n of Nerode equivalence classes is the number of distinct left quotients of L , known as its *quotient complexity* [1]. This is the same number as the number of states in \mathcal{D} , and is therefore known as L 's *state complexity* [16]. Quotient/state complexity is now a commonly used measure of complexity of a regular language, and constitutes a basic reference for other measures of complexity. One can also define the quotient complexity of a Nerode equivalence class, that is, of the language accepted by DFA ${}_q\mathcal{D}$. In the worst case – for example, if \mathcal{D} is strongly connected – this is n for every q .

The *Myhill congruence* [13] refines the Nerode right congruence and is a (two-sided) congruence. Here a word x is equivalent to a word y if for all u and v in Σ^* , uxv is in L if and only if uyv is in L . This is also known as the *syntactic congruence* [15] of L . The quotient of the free semigroup Σ^+ by this congruence is the *syntactic semigroup* of L . In automaton-theoretic terms two words are equivalent if they induce the same transformation of the set of states of a minimal DFA of L . The quotient complexity of Myhill classes has not been studied.

The third equivalence, which we call the *atom congruence* is a left congruence refined by the Myhill congruence. Here 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 called an *atom* of L [9]. It follows that an atom is a non-empty intersection of complemented and uncomplemented quotients of L .

The atom congruence is related to the Myhill and Nerode congruences in a natural way. Say a congruence on Σ^* *recognizes* L if L can be written as a union of the congruence's classes. The Myhill congruence is the unique *coarsest* congruence (that is, the one with the fewest equivalence classes) that recognizes L [15]. The Nerode and atom congruences are respectively the coarsest *right* and *left* congruences that recognize L .

The quotient complexity of atoms of regular languages has been studied in [4, 8, 12]. In this paper we study the quotient complexity of atoms in three subclasses of regular languages, namely, right, left, and two-sided ideals.

Ideals are fundamental concepts in semigroup theory. A language L over an alphabet Σ is a *right* (respectively, *left* and *two-sided*) *ideal* if $L = L\Sigma^*$ (respectively, $L = \Sigma^*L$ and $L = \Sigma^*L\Sigma^*$). The quotient complexity of operations on regular ideal languages has been studied in [6], and the reader should refer to that paper for more information about ideals. Ideals appear in pattern matching. A right (left)

ideal $L\Sigma^*$ (Σ^*L) represents the set of all words beginning (ending) with some word of a given set L , and $\Sigma^*L\Sigma^*$ is the set of all words containing a factor from L .

2 Preliminaries

It is well known that a language $L \subseteq \Sigma^*$ is regular if and only if it has a finite number of quotients. We denote the number of quotients of L (the *quotient complexity*) by $\kappa(L)$. This is the same as the *state complexity*, the number of states in a minimal DFA of L . Since we will not be discussing other measures of complexity, we refer to both quotient and state complexity as just *complexity*.

Let the set of quotients of a regular language L be $K = \{K_1, \dots, K_n\}$. The *quotient automaton* of L is the DFA $\mathcal{D} = (K, \Sigma, \delta, L, F)$, where $\delta(K_i, a) = K_j$ if $a^{-1}K_i = K_j$, $L = K_1 = \varepsilon^{-1}L$ by convention, and $F = \{K_i \mid \varepsilon \in K_i\}$. This DFA is uniquely defined by L and is isomorphic to every minimal DFA of L .

A *transformation* of a set Q_n of n elements is a mapping of Q_n into itself, whereas a *permutation* of Q_n is a mapping of Q_n onto itself. In this paper we consider only transformations of finite sets, and we assume without loss of generality that $Q_n = \{1, \dots, n\}$. An arbitrary transformation has the form

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

where $i_k \in Q_n$ for $1 \leq k \leq n$. The image i_j of element j under transformation t is denoted by jt . The image of $S \subseteq Q_n$ is $St = \bigcup_{j \in S} \{jt\}$. The *identity* transformation $\mathbf{1}$ maps each element to itself. For $k \geq 2$, a transformation (permutation) t is a *k-cycle* if there is a set $P = \{q_1, q_2, \dots, q_k\} \subseteq Q_n$ such that if $q_1t = q_2, q_2t = q_3, \dots, q_{k-1}t = q_k, q_kt = q_1$, and $qt = q$ for all $q \notin P$. A *k-cycle* is denoted by (q_1, q_2, \dots, q_k) . A 2-cycle (q_1, q_2) is called a *transposition*. A transformation is *constant* if it maps all states to a single state q ; we denote it by $(Q_n \rightarrow q)$. A transformation t that maps p to q , $q \neq p$ and does not affect any $r \neq p$ is denoted by $(p \rightarrow q)$. The set of all transformations of Q_n is a monoid under composition, called the *complete transformation monoid* and denoted by T_n . The following is well-known:

Proposition 1. *The complete transformation monoid T_n has size n^n and can be generated by $\{(1, \dots, n), (1, 2), (n \rightarrow 1)\}$, and by $\{(1, \dots, n), (2, \dots, n), (n \rightarrow 1)\}$.*

For a DFA $\mathcal{D} = (Q, \Sigma, \delta, q_1, F)$ we define the transformations $\{\delta_w \mid w \in \Sigma^+\}$ by $q\delta_a = \delta(q, a)$ for $a \in \Sigma$, and $q\delta_w = q\delta_x\delta_a$ for $w = xa$, $x \in \Sigma^*$. This set is a semigroup under composition and it is called the *transition semigroup* of \mathcal{D} . We also define $\delta_\varepsilon = \mathbf{1}$. The transformation δ_w is called the *transformation induced by w* . To simplify notation, we usually make no distinction between the word $w \in \Sigma^*$ and the transformation δ_w . If \mathcal{D} is the quotient automaton of L , then the transition semigroup of \mathcal{D} is isomorphic to the syntactic semigroup of L [15]. A state $q \in Q$ is *reachable from $p \in Q$* if $pw = q$ for some $w \in \Sigma^*$, and *reachable* if it is reachable from q_1 . Two states p, q are *indistinguishable* if $pw \in F \Leftrightarrow qw \in F$ for all $w \in \Sigma^*$,

and *distinguishable* otherwise. Indistinguishability is an equivalence relation on Q ; furthermore, if \mathcal{D} recognizes a language L , we can compute $\kappa(L)$ by counting the number of equivalence classes under indistinguishability of the reachable states of \mathcal{D} . A state is *empty* if its right language (defined in the introduction) is \emptyset .

3 Atoms

Atoms of regular languages were studied in [9], and their complexities in [3, 8, 12]. As discussed earlier, atoms are the classes of the *atom congruence*, a left congruence which is the natural counterpart of the Myhill two-sided congruence and Nerode right congruence. The Myhill and Nerode congruences are fundamental in regular language theory, but it seems comparatively little attention has been paid to the atom congruence and its classes. In [2] it was argued that it is useful to consider the complexity of a language's atoms when searching for complex regular languages, since one would expect such languages to have complex atoms.

Below we present an alternative characterization of atoms, which we use in our proofs. Earlier papers on atoms such as [3, 8, 9] take this as the definition of atoms, for it was not known until recently that atoms may be viewed as congruence classes (this fact was first noticed by Iván in [12]).

From now on assume all languages are non-empty. Denote the complement of a language L by $\bar{L} = \Sigma^* \setminus L$. Let $Q_n = \{1, \dots, n\}$ and let L be a regular language with quotients $K = \{K_1, \dots, K_n\}$. Each subset S of Q_n defines an *atomic intersection* $A_S = \bigcap_{i \in S} K_i \cap \bigcap_{i \in \bar{S}} \bar{K}_i$, where $\bar{S} = Q_n \setminus S$. An *atom* of L is a non-empty atomic intersection. Since atoms are pairwise disjoint, every atom A has a unique atomic intersection associated with it, and this atomic intersection has a unique subset S of K associated with it. This set S is called the *basis* of A .

Throughout the paper, L is a regular language of complexity n with quotients K_1, \dots, K_n and minimal DFA $\mathcal{D} = (Q_n, \Sigma, \delta, 1, F)$ such that the language of state i is K_i . Let $A_S = \bigcap_{i \in S} K_i \cap \bigcap_{i \in \bar{S}} \bar{K}_i$ be an atom. For any $w \in \Sigma^*$ we have

$$w^{-1}A_S = \bigcap_{i \in S} w^{-1}K_i \cap \bigcap_{i \in \bar{S}} \overline{w^{-1}K_i}.$$

Since a quotient of a quotient of L is also a quotient of L , $w^{-1}A_S$ has the form;

$$w^{-1}A_S = \bigcap_{i \in X} K_i \cap \bigcap_{i \in Y} \bar{K}_i,$$

where $|X| \leq |S|$ and $|Y| \leq n - |S|$, $X, Y \subseteq Q_n$.

The complexity of atoms of a regular language was computed in [8] using a unique NFA defined by L_n , called the *átomaton*. In that NFA the language of each state q_S is an atom A_S of L_n . To find the complexity of that atom, the *átomaton* started in state q_S was converted to an equivalent DFA. A more direct and simpler method was used by Szabolcs Iván [12] who constructed the DFA for the atom directly from the DFA \mathcal{D}_n . We follow that approach here and outline it briefly for completeness.

For any regular language L an atom A_S corresponds to the ordered pair (S, \overline{S}) , where S (\overline{S}) is the set of subscripts of uncomplemented (complemented) quotients. If L is represented by a DFA $\mathcal{D} = (Q, \Sigma, \delta, q_1, F)$, it is more convenient to think of S and \overline{S} as subsets of Q . Similarly, any quotient of A_S corresponds to a pair (X, Y) of subsets of Q . For the quotient of A_S reached when a letter $a \in \Sigma$ is applied to the quotient corresponding to (X, Y) we get

$$a^{-1} \left(\bigcap_{i \in X} K_i \cap \bigcap_{i \in Y} \overline{K_i} \right) = \bigcap_{i \in X} a^{-1}K_i \cap \bigcap_{i \in Y} \overline{a^{-1}K_i} = \bigcap_{i \in X} K_{ia} \cap \bigcap_{i \in Y} \overline{K_{ia}}.$$

In terms of pairs of subsets of Q , from (X, Y) we reach (Xa, Ya) . Note that if $X \cap Y \neq \emptyset$ in (X, Y) then the corresponding quotient is empty. Note also that the quotient of atom A_S corresponding to (X, Y) is final if and only if each quotient K_i with $i \in X$ contains ε , and each K_j with $j \in Y$ does not contain ε .

These considerations lead to the following definition of a DFA for A_S [12]:

Definition 1. Suppose $\mathcal{D} = (Q, \Sigma, \delta, q_1, F)$ is a DFA and let $S \subseteq Q$. Define the DFA $\mathcal{D}_S = (Q_S, \Sigma, \Delta, (S, \overline{S}), F_S)$, where

- $Q_S = \{(X, Y) \mid X, Y \subseteq Q, X \cap Y = \emptyset\} \cup \{\perp\}$.
- For all $a \in \Sigma$, $\Delta((X, Y), a) = (\delta(X, a), \delta(Y, a))$ if $\delta(X, a) \cap \delta(Y, a) = \emptyset$, and $\Delta((X, Y), a) = \perp$ otherwise; and $\Delta(\perp, a) = \perp$.
- $F_S = \{(X, Y) \mid X \subseteq F, Y \subseteq \overline{F}\}$.

DFA \mathcal{D}_S recognizes the atomic intersection A_S of L . If \mathcal{D}_S recognizes a non-empty language, then A_S is an atom.

4 Complexity of Atoms in Regular Languages

Upper bounds on the maximal complexity of atoms of regular languages were derived in [8]; for completeness we include these results. For $n = 1$ there is only one non-empty language $L = \Sigma^*$; it has one atom, L , which is of complexity 1. From now on assume that $n \geq 2$.

Proposition 2. Let L be a regular language with $n \geq 2$ quotients. Then L has at most 2^n atoms. If $S \in \{Q_n, \emptyset\}$, then $\kappa(A_S) \leq 2^n - 1$. Otherwise,

$$\kappa(A_S) \leq 1 + \sum_{x=1}^{|S|} \sum_{y=1}^{n-|S|} \binom{n}{x} \binom{n-x}{y}.$$

Proof. Since the number of subsets S of Q_n is 2^n , there are at most that many atoms. For atom complexity consider the following three cases:

1. $S = Q_n$. Then $A_{Q_n} = \bigcap_{i \in Q_n} K_i$ is the intersection of all quotients of L . For $w \in \Sigma^*$, we have $w^{-1}A_{Q_n} = \bigcap_{i \in X} K_i$, where $1 \leq |X| \leq |Q_n|$. Hence there are at most $2^n - 1$ quotients of this atom.

2. $S = \emptyset$. Now $A_\emptyset = \bigcap_{i \in Q_n} \overline{K_i}$, and $w^{-1}A_\emptyset = \bigcap_{i \in Y} \overline{K_i}$, where $1 \leq |Y| \leq |Q_n|$. As in the first case, there are at most $2^n - 1$ quotients of this atom.
3. $\emptyset \subsetneq S \subsetneq Q_n$. Then $A_S = \bigcap_{i \in S} K_i \cap \bigcap_{i \in \overline{S}} \overline{K_i}$. Every quotient of A_S has the form $w^{-1}A_S = \bigcap_{i \in X} K_i \cap \bigcap_{i \in Y} \overline{K_i}$, where $1 \leq |X| \leq |S|$ and $1 \leq |Y| \leq n - |S|$. There are two subcases:
 - a) If $X \cap Y \neq \emptyset$, then $w^{-1}A_S = \emptyset$.
 - b) If $X \cap Y = \emptyset$, there are at most $\sum_{x=1}^{|S|} \sum_{y=1}^{n-|S|} \binom{n}{x} \binom{n-x}{y}$ quotients of A_S of this form. This follows since $\binom{n}{x}$ is the number of ways to choose a set $X \subseteq Q_n$ of size x , and once X is fixed, $\binom{n-x}{y}$ is the number of ways to choose a set $Y \subseteq Q_n$ of size y that is *disjoint* from X . Taking the sum over the permissible values of x and y gives the formula above.

Adding the results of (a) and (b) we have the required bound.

□

It was shown in [2] that the language L_n accepted by the minimal DFA \mathcal{D}_n of Definition 2, also illustrated in Figure 1, meets all the complexity bounds for common operations on regular languages.

Definition 2. For $n \geq 2$, let $\mathcal{D}_n = (Q_n, \Sigma, \delta_n, 1, \{n\})$, where $Q_n = \{1, \dots, n\}$ is the set of states, $\Sigma = \{a, b, c\}$ is the alphabet, the transition function δ_n is defined by $a = (1, \dots, n)$, $b = (1, 2)$, and $c = (n \rightarrow 1)$, state 1 is the initial state, and $\{n\}$ is the set of final states. Let L_n be the language accepted by \mathcal{D}_n . (If $n = 2$, a and b induce the same transformation; hence $\Sigma = \{a, c\}$ suffices.)

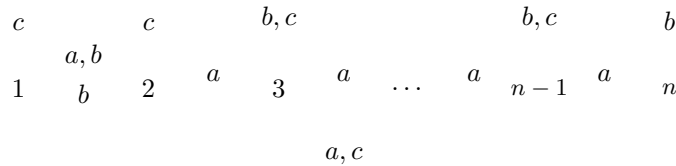


Figure 1: DFA of a regular language whose atoms meet the bounds.

It was proved in [8] that L_n has 2^n atoms, all of which are as complex as possible. We include the proof of this theorem following [12]. We first prove a general result about distinguishability of states in \mathcal{D}_S , which we will use throughout the paper.

Lemma 1 (Distinguishability). Let $\mathcal{D} = (Q, \Sigma, \delta, q_1, F)$ be a minimal DFA and for $S \subseteq Q$, let $\mathcal{D}_S = (Q_S, \Sigma, \Delta, (S, \overline{S}), F_S)$ be the DFA of the atom A_S . Then:

1. States (X, Y) and (X', Y') of \mathcal{D}_S are distinguishable if $X \neq X'$ and $A_X, A_{X'}$ are both atoms, or if $Y \neq Y'$ and $A_{\overline{Y}}, A_{\overline{Y}'}$ are both atoms.
2. If one of A_X or $A_{\overline{Y}}$ is an atom, then (X, Y) is distinguishable from \perp .

Proof. First note that if A_Z is an atom, then the initial state of \mathcal{D}_Z must be non-empty, so there is a word w_Z in the transition semigroup of \mathcal{D} such that $(Z, \overline{Z})w_Z = (U, V)$ with $U \subseteq F, V \subseteq \overline{F}$, i.e., $(U, V) \in F_S$. In particular, $(X, Y)w_X \in F_S$, since $Y \subseteq \overline{X}$. We also have $(X, Y)w_{\overline{Y}} \in F_S$, since Y is sent to a subset of \overline{F} , and $X \subseteq \overline{Y}$ is sent to a subset of F . This proves (2): if one of A_X or $A_{\overline{Y}}$ is an atom, then one of w_X or $w_{\overline{Y}}$ is in the transition semigroup of \mathcal{D} , and hence (X, Y) can be mapped to a final state but \perp cannot. Now, we consider the two cases from (1):

1. $X \neq X'$. Suppose $X' \not\subseteq X$. Then $(X, Y)w_X \in F_S$, but $(X', Y')w_X \notin F_S$, since $X' \setminus X$ is a non-empty subset of \overline{X} and hence gets mapped outside of F . Thus w_X distinguishes these states. If instead we have $X \not\subseteq X'$, then $w_{X'}$ distinguishes the states. Hence if $A_X, A_{X'}$ are atoms, w_X and $w_{X'}$ are in the transition semigroup of \mathcal{D} , and the states are distinguishable.
2. $Y \neq Y'$. If $Y' \not\subseteq Y$, then $w_{\overline{Y}}$ distinguishes (X, Y) from (X', Y') ; otherwise, $w_{\overline{Y}'}$ distinguishes the states. As before, if $A_{\overline{Y}}, A_{\overline{Y}'}$ are atoms then the states are distinguishable.

□

Theorem 1. For $n \geq 2$, the language L_n of Definition 2 has 2^n atoms and each atom meets the bounds of Proposition 2.

Proof. The DFA for the atomic intersection A_S is $\mathcal{D}_S = (Q_S, \Sigma, \Delta, (S, \overline{S}), F_S)$, where $F_S = \{(X, Y) \mid X \subseteq \{n\}, Y \subseteq Q_n \setminus \{n\}\}$. By Proposition 1, the transition semigroup of \mathcal{D}_n consists of all n^n transformations of the state set Q_n . Hence (S, \overline{S}) can be mapped to a final state in F_S by a transformation that sends S to $\{n\}$ and \overline{S} to $\{1\}$. It follows that all 2^n atomic intersections $A_S, S \subseteq Q_n$ are atoms. By the Distinguishability Lemma, all distinct states in \mathcal{D}_S are distinguishable. It suffices to prove the number of reachable states in each \mathcal{D}_S meets the bounds.

If $S = Q_n$, then the DFA \mathcal{D}_S of A_S has initial state (Q_n, \emptyset) . The reachable states of \mathcal{D}_S are of the form (X, \emptyset) , where X is the image of Q_n under some transformation in the transition semigroup. Since we have all transformations, we can reach all $2^n - 1$ states $(X, \emptyset), \emptyset \subsetneq X \subseteq Q_n$. For $S = \emptyset$ a similar argument works.

If $\emptyset \subsetneq S \subsetneq Q_n$, then for any state (X, Y) with $1 \leq X \leq |S|, 1 \leq Y \leq n - |S|$ and $X \cap Y = \emptyset$, we can find a transformation mapping S onto X and \overline{S} onto Y . So all these states are reachable, and there are $\sum_{x=1}^{|S|} \sum_{y=1}^{n-|S|} \binom{n}{x} \binom{n-x}{y}$ of them. In addition, \perp is reachable from (S, \overline{S}) by the constant transformation $(Q_n \rightarrow 1)$ and so the bound is met. □

5 Complexity of Atoms in Right Ideals

If L is a right ideal, one of its quotients is Σ^* ; by convention we assume that $K_n = \Sigma^*$. In any atom A_S the quotient K_n must be uncomplemented, that is, we must have $n \in S$. Thus A_\emptyset is not an atom. The results of this section were stated in [4] without proof; for completeness we include the proofs.

Proposition 3. *Suppose L is a right ideal with $n \geq 1$ quotients. Then L has at most 2^{n-1} atoms. The complexity $\kappa(A_S)$ of atom A_S satisfies*

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

Proof. Let A_S be an atom. Since $w^{-1}\Sigma^* = \Sigma^*$ for all $w \in \Sigma^*$, $w^{-1}A_S$ always has K_n uncomplemented; so if (X, Y) corresponds to $w^{-1}A_S$, then $n \in X$. Since the number of subsets S of Q_n containing n is 2^{n-1} , there are at most that many atoms. Consider two cases:

1. $S = Q_n$. Then $w^{-1}L = \bigcap_{i \in X} K_i$, and each such quotient of A_S is represented by (X, \emptyset) , where $1 \leq |X| \leq n$. Since n is always in X , there are at most 2^{n-1} quotients of this atom.
2. $\emptyset \subsetneq S \subsetneq Q_n$. Then $w^{-1}A_S = \bigcap_{i \in X} K_i \cap \bigcap_{i \in Y} \overline{K_i}$, where $1 \leq |X| \leq |S|$ and $1 \leq |Y| \leq n - |S|$. We know that if $X \cap Y \neq \emptyset$, then $w^{-1}A_S = \emptyset$. Thus we are looking for pairs (X, Y) such that $n \in X$ and $X \cap Y = \emptyset$. To get X we take n and choose $|X| - 1$ elements from $Q_n \setminus \{n\}$, and then to get Y we take $|Y|$ elements from $Q_n \setminus X$. The number of such pairs is $\sum_{x=1}^{|S|} \sum_{y=1}^{n-|S|} \binom{n-1}{x-1} \binom{n-x}{y}$. Adding the empty quotient we have our bound.

□

For $n = 1$, $L = \Sigma^*$ is a right ideal with one atom of complexity 1. For $n = 2$, $L = aa^*$ is a right ideal with two atoms L and \overline{L} of complexity 2. It was shown in [4] that the languages of the DFAs of Definition 3 are “most complex” amongst right ideals, in the sense that they meet all the complexity bounds for common operations, but no proof of atom complexity was given. We include this proof here.

Definition 3. *For $n \geq 3$, let $\mathcal{D}_n = (Q_n, \Sigma, \delta_n, 1, \{n\})$, where $\Sigma = \{a, b, c, d\}$, and δ_n is defined by $a = (1, \dots, n-1)$, $b = (2, \dots, n-1)$, $c = (n-1 \rightarrow 1)$ and $d = (n-1 \rightarrow n)$. Let L_n be the language accepted by \mathcal{D}_n . If $n = 3$, b is not needed; hence $\Sigma = \{a, c, d\}$ suffices. Also, let $L_2 = aa^*$ and $L_1 = a^*$.*

Theorem 2. *For $n \geq 1$, the language L_n of Definition 3 is a right ideal that has 2^{n-1} atoms and each atom meets the bounds of Proposition 3.*

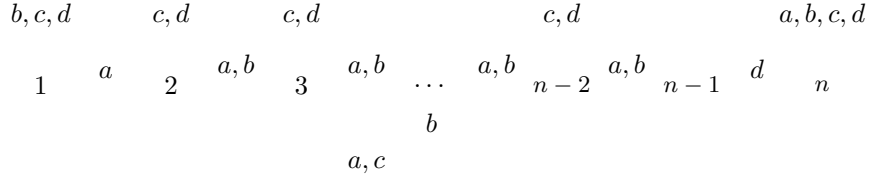


Figure 2: DFA of a right ideal whose atoms meet the bounds.

Proof. The cases $n < 3$ are easily verified; hence assume $n \geq 3$. By Proposition 1, the transformations $\{a, b, c\}$ restricted to Q_{n-1} generate all transformations of Q_{n-1} . When d is included, we get all transformations of Q_n that fix n . For $S \subseteq Q_n$, $n \in S$, consider the DFA \mathcal{D}_S , which has initial state (S, \bar{S}) . There is a transformation of Q_n fixing n that sends (S, \bar{S}) to the final state $(\{n\}, \{1\})$. Hence A_S is an atom if $n \in S$, and so L_n has 2^{n-1} atoms.

We now count reachable and distinguishable states in the DFA of each atom. Suppose $S = Q_n$. The initial state of \mathcal{D}_S is (Q_n, \emptyset) ; by transformations that fix n , we can reach any state (X, \emptyset) with $\{n\} \subseteq X \subseteq Q_n$. There are 2^{n-1} such states, and since A_X is an atom if $n \in X$, all of them are distinguishable by the Distinguishability Lemma.

Suppose $\emptyset \subsetneq S \subsetneq Q_n$. From the initial state (S, \bar{S}) , by transformations that fix n we can reach any (X, Y) with $1 \leq |X| \leq |S|$, $1 \leq |Y| \leq n - |S|$, $n \in X$ and $X \cap Y = \emptyset$. There are $\sum_{x=1}^{|S|} \sum_{y=1}^{n-|S|} \binom{n-1}{x-1} \binom{n-x}{y}$ such states. For all such states (X, Y) , we have $n \in X$ and $n \in \bar{Y}$, so A_X and $A_{\bar{Y}}$ are both atoms; hence by the Distinguishability Lemma, all of these states are distinguishable from each other and from \perp . The state \perp is also reachable by the constant transformation $(Q_n \rightarrow n)$, and so the bound is met. □

6 Complexity of Atoms in Left Ideals

If L is a left ideal, then $L = \Sigma^*L$, and $w^{-1}L$ contains L for every $w \in \Sigma^*$. By convention we let $L = K_1$.

Proposition 4. *Suppose L is a left ideal with $n \geq 2$ quotients. Then L has at most $2^{n-1} + 1$ atoms. The complexity $\kappa(A_S)$ of atom A_S satisfies*

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

Proof. Consider the atomic intersections A_S such that $1 \in S$; then $\bigcap_{i \in S} K_i = L$ (since every quotient contains L), and there are two possibilities: Either $S = Q_n$,

in which case $A_S = A_{Q_n} = \bigcap_{i \in Q_n} K_i = L$, or there is at least one quotient, say K_i which is complemented. Since K_i contains L , it can be expressed as $K_i = L \cup M_i$, where $L \cap M_i = \emptyset$. Then the intersection has the term $L \cap \overline{(L \cup M_i)} = \emptyset$, and A_S is not an atom. Thus for A_S to be an atom, either $1 \notin S$ or $S = Q_n$. Hence there are at most $2^{n-1} + 1$ atoms.

For atom complexity, consider the following cases:

1. $S = Q_n$. Then $A_{Q_n} = L$, and the complexity of A_{Q_n} is precisely n .
2. $S = \emptyset$. Now $A_\emptyset = \bigcap_{i \in Q_n} \overline{K_i}$, and every quotient of A_\emptyset is an intersection $\bigcap_{i \in Y} \overline{K_i}$, where $1 \leq |Y| \leq |Q_n|$. There are $2^n - 1$ such intersections, but consider any quotient $K_i \neq L$ of a left ideal; it can be expressed as $K_i = L \cup M_i$, where $L \cap M_i = \emptyset$. We have

$$\overline{K_1} \cap \overline{K_i} = \overline{L \cup M_i} = \overline{L} \cap \overline{L \cup M_i} = \overline{L} \cap \overline{M_i} = \overline{K_i}.$$

Thus every intersection $\bigcap_{i \in Y} \overline{K_i}$ which has $Y \neq \emptyset$ and does not have $\overline{K_1}$ as a term defines the same language as $\overline{K_1} \cap \bigcap_{i \in Y} \overline{K_i}$. There are $2^{n-1} - 1$ such intersections. Adding 1 for the intersection which just has the single term $\overline{K_1}$, we get our bound 2^{n-1} .

3. $\emptyset \subsetneq S \subsetneq Q_n$. Then $A_S = \bigcap_{i \in S} K_i \cap \bigcap_{i \in \overline{S}} \overline{K_i}$, where neither S nor \overline{S} is empty. If $1 \in S$ then A_S is not an atom, so assume $1 \notin S$. Every quotient of A_S has the form $w^{-1}A_S = \bigcap_{i \in X} K_i \cap \bigcap_{i \in Y} \overline{K_i}$, where $1 \leq |X| \leq |S|$ and $1 \leq |Y| \leq n - |S|$.

a) $1 \in X$. We claim that $w^{-1}A_S = \emptyset$ for all $w \in \Sigma^*$. For suppose that there is a term K_i , $i \in S$, and a word $w \in \Sigma^*$ such that $w^{-1}K_i = K_1$. Since $K_1 \subseteq K_i$, we have $w^{-1}K_1 \subseteq w^{-1}K_i = K_1$. Since also $K_1 \subseteq w^{-1}K_1$ because L is a left ideal, we have $w^{-1}K_1 = K_1$. But $1 \in \overline{S}$, so $w^{-1}(\bigcap_{i \in \overline{S}} \overline{K_i}) = \bigcap_{i \in Y} \overline{K_i}$ has $w^{-1}\overline{K_1} = \overline{K_1}$ as a term. Thus $1 \in Y$, which means $X \cap Y \neq \emptyset$. Hence $w^{-1}A_S = \emptyset$.

b) $1 \notin X$. We are looking for pairs (X, Y) such that $X \cap Y = \emptyset$. As we argued in (2), $\overline{K_1} \cap \overline{K_i} = \overline{K_i}$ for each i , so we can assume without loss of generality that $1 \in Y$. To get X we choose $|X|$ elements from $Q_n \setminus \{1\}$ and to get Y we take $\{1\}$ and choose $|Y| - 1$ elements from $(Q_n \setminus X) \setminus \{1\}$. The number of such pairs is $\sum_{x=1}^{|S|} \sum_{y=1}^{n-|S|} \binom{n-1}{x} \binom{n-x-1}{y-1}$.

Adding 1 for the empty quotient we have our bound.

□

Next we compare the bounds for left ideals with those for right ideals. To calculate the number of pairs (X, Y) such that $n \in X$ and $X \cap Y = \emptyset$ for right

ideals, we can first choose Y from $Q_n \setminus \{n\}$, and then choose X by taking n and choosing $|X| - 1$ elements from $(Q_n \setminus Y) \setminus \{n\}$. The number of such pairs is

$$1 + \sum_{y=1}^{n-|S|} \sum_{x=1}^{|S|} \binom{n-1}{y} \binom{n-y-1}{x-1}.$$

If we interchange x and y we note that this is precisely the number of pairs (X, Y) such that $1 \in Y$ and $X \cap Y = \emptyset$ for an atom of a left ideal with a basis of size $n - |S|$. Thus we have

Remark 1. Let R be a right ideal of complexity n and let A_S be an atom of R , where $\emptyset \subsetneq S \subsetneq Q_n$. Let L be a left ideal of complexity n and let A'_S be an atom of L . The upper bounds on the complexities of A_S and A'_S are equal.

Now we consider the question of tightness of the bounds in Proposition 4. For $n = 1$, $L = \Sigma^*$ is a left ideal with one atom of complexity 1; so the bound of Proposition 4 does not hold.

The DFAs of Definition 4 and Figure 3 were introduced in [10]. It was shown in [7] that the languages of these DFAs have the largest syntactic semigroups amongst left ideals of complexity n . Moreover, it was shown in [5] that these languages also meet the bounds on the quotient complexity of boolean operations, concatenation and star. Together with our result about the number of atoms and their complexity, this shows that these languages are “most complex” left ideals.

Definition 4. For $n \geq 3$, let $\mathcal{D}_n = (Q_n, \Sigma, \delta_n, 1, \{n\})$, where $\Sigma = \{a, b, c, d, e\}$, and δ_n is defined by $a = (2, \dots, n)$, $b = (2, 3)$, $c = (n \rightarrow 2)$, $d = (n \rightarrow 1)$, and $e = (Q_n \rightarrow 2)$. If $n = 3$, inputs a and b coincide; hence $\Sigma = \{a, c, d, e\}$ suffices. Also, let $\mathcal{D}_2 = (Q_2, \{a, b, c\}, \delta_2, 1, \{2\})$, where $a = \mathbf{1}$, $b = (Q_2 \rightarrow 2)$, $c = (Q_2 \rightarrow 1)$. Let L_n be the language accepted by \mathcal{D}_n ; we have $L_2 = \Sigma^*b(a \cup b)^*$.

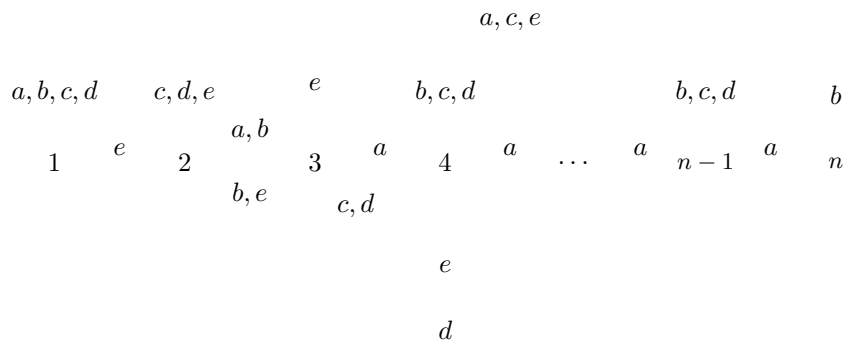


Figure 3: DFA of a left ideal whose atoms meet the bounds.

Theorem 3. For $n \geq 2$, the language L_n of Definition 4 is a left ideal that has $2^{n-1} + 1$ atoms and each atom meets the bounds of Proposition 4.

Proof. It was proved in [10] that L_n is a left ideal of complexity n . The case $n = 2$ is easily verified; hence assume $n \geq 3$. It was proved in [7] that the transition semigroup of \mathcal{D}_n contains all transformations of Q_n that fix 1 and all constant transformations. Recall that if A_S is an atom of a left ideal, then either $S = Q_n$ or $1 \notin S$. For all S with $1 \notin S$, from (S, \bar{S}) we can reach the final state $(\{n\}, \{1\})$ of \mathcal{D}_S (or $(\emptyset, \{1\})$ for $S = \emptyset$) by transformations that fix 1. For $S = Q_n$, let $w = (Q_n \rightarrow n)$; then $(Q_n, \emptyset)w = (\{n\}, \emptyset)$ is final in \mathcal{D}_S . Hence if $S = Q_n$ or $1 \notin S$, then A_S is an atom of L_n , and so L has $2^{n-1} + 1$ atoms.

We now count reachable and distinguishable states in the DFA of each atom. We know that A_{Q_n} has complexity n for all left ideals, so assume $1 \notin S$. If $S = \emptyset$, the initial state of \mathcal{D}_S is (\emptyset, Q_n) . By transformations that fix 1 we can reach (\emptyset, Y) for all Y with $\{1\} \subseteq Y \subseteq Q_n$. There are 2^{n-1} of these states. Since \bar{Y} does not contain 1, $A_{\bar{Y}}$ is an atom, so all of these states are distinguishable by the Distinguishability Lemma.

If $\emptyset \subsetneq S \subsetneq Q_n$, the initial state of \mathcal{D}_S is (S, \bar{S}) . Since $1 \notin S$, by transformations that fix 1, we can reach any state (X, Y) with $1 \leq |X| \leq |S|$, $1 \leq |Y| \leq n - |S|$, $1 \notin X$, $1 \in Y$, and $X \cap Y = \emptyset$. There are $\sum_{x=1}^{|S|} \sum_{y=1}^{n-|S|} \binom{n-1}{x} \binom{n-x-1}{y-1}$ such states. They are all distinguishable from each other and from \perp by the Distinguishability Lemma, since $1 \notin X$, $1 \in Y$ imply that A_X and $A_{\bar{Y}}$ are both atoms. We can also reach \perp from (S, \bar{S}) by any constant transformation, and so the bound is met. \square

7 Complexity of Atoms in Two-Sided Ideals

7.1 Upper Bounds

A language is a two-sided ideal if it is both a right ideal and a left ideal.

Proposition 5. Suppose L is a two-sided ideal with $n \geq 2$ quotients. Then L has at most $2^{n-2} + 1$ atoms. The complexity $\kappa(A_S)$ of atom A_S satisfies

$$\kappa(A_S) \begin{cases} = n, & \text{if } S = Q_n; \\ \leq 2^{n-2} + n - 1, & \text{if } S = Q_n \setminus \{1\}; \\ \leq 1 + \sum_{x=1}^{|S|} \sum_{y=1}^{n-|S|} \binom{n-2}{x-1} \binom{n-x-1}{y-1}, & \text{otherwise.} \end{cases} \quad (3)$$

Proof. Since L is a left ideal, A_S is an atom only if $S = Q_n$ or $S \subseteq Q_n \setminus \{1\}$; since L is a right ideal we must also have $n \in S$. This gives our upper bound of $2^{n-2} + 1$ atoms.

We know that A_{Q_n} has complexity n since this is true for left ideals. Since L is a right ideal, A_\emptyset is not an atom, so we can assume $S \neq \emptyset$.

Suppose A_S is an atom of L , with $S \neq Q_n$ and $S \neq Q_n \setminus \{1\}$. We proved for left ideals that the number of distinct non-empty quotients of A_S is bounded by the number of pairs (X, Y) , $1 \leq |X| \leq |S|$, $1 \leq |Y| \leq n - |S|$, $1 \notin X$, $1 \in Y$, $X \cap Y = \emptyset$.

Since L is a right ideal, we must also have $n \in X$ and $n \notin Y$. There are $\binom{n-2}{|X|-1}$ possibilities for X , since X must contain n and the remaining $|X| - 1$ elements are taken from $Q_n \setminus \{1, n\}$. If X is fixed, there are $\binom{n-|X|-1}{|Y|-1}$ possibilities for Y , since Y must contain 1 and the remaining $|Y| - 1$ elements are taken from $(Q_n \setminus X) \setminus \{1\}$. Since $Q_n \setminus X$ always contains 1, the size of $(Q_n \setminus X) \setminus \{1\}$ is always $n - |X| - 1$. Summing over the possible sizes of X and Y and adding 1 for the empty quotient, we get the required bound.

This leaves the case of $S = Q_n \setminus \{1\}$. Each quotient of A_S has the form

$$w^{-1}A_S = \left(\bigcap_{i \in X} K_i \right) \cap \overline{K_j},$$

where $K_j = w^{-1}K_1 = w^{-1}L$, and $n \in X$. We can view the non-empty quotients as states $(X, \{j\})$ of the DFA \mathcal{D}_S for A_S , where \mathcal{D} is a minimal DFA for L . We must have $n \in X$ and $X \cap \{j\} = \emptyset$, and so $j \notin X$.

For each p in Q_n , define the set $S(p) = \{q \in Q_n \mid K_p \subsetneq K_q\}$. The elements of $S(p)$ are called the *successors* of p . Note that p is not a successor of itself. We claim that if the quotient $w^{-1}A_S$ is non-empty and the corresponding state of \mathcal{D}_S is $(X, \{j\})$, then $X \subseteq S(j)$.

To see this, note that since L is a left ideal, we have $L \subseteq K_i$ for all $i \in Q_n$. It follows that $w^{-1}L = K_j \subseteq w^{-1}K_i$ for all $i \in Q_n$. Thus in the formula for $w^{-1}A_S$ above, we have $K_j \subseteq K_i$ for all $i \in X$. But if $K_j = K_i$ for any $i \in X$, then $w^{-1}A_S$ is empty. Thus $K_j \subsetneq K_i$ for all $i \in X$, which implies that $X \subseteq S(j)$.

X must contain n , since L is a right ideal. Thus for each j , there are at most $2^{|S(j)|-1}$ states $(X, \{j\})$. The index j can range from 1 to $n - 1$; we cannot have $j = n$ since $n \in X$ but $j \notin X$. This gives an upper bound of $\sum_{j=1}^{n-1} 2^{|S(j)|-1}$ for the number of non-empty quotients.

This bound is not tight, so we refine it by considering the distinguishability relations between states of \mathcal{D}_S . Choose $i \neq n \in S(j)$ and a non-empty set $Y \subseteq S(i) \setminus \{n\}$. Then $K_i \subsetneq K_q$ for all $q \in Y$, so we have $K_i \cap \left(\bigcap_{q \in Y} K_q \right) = K_i$. This means $(\{i, n\}, \{j\})$ is indistinguishable from $(Y \cup \{i, n\}, \{j\})$. Since Y is non-empty and does not contain n , there are at most $2^{|S(i)|-1} - 1$ possibilities for Y .

From this we get a new upper bound for the number of distinguishable states $(X, \{j\})$ for a fixed j , as follows: first take our previous bound of $2^{|S(j)|-1}$. Then for each $i \neq n \in S(j)$, subtract $2^{|S(i)|-1} - 1$ to account for the states $(Y \cup \{i, n\}, \{j\})$ that are equivalent to $(\{i, n\}, \{j\})$. Our new bound is

$$2^{|S(j)|-1} - \sum_{\substack{i \in S(j) \\ i \neq n}} (2^{|S(i)|-1} - 1) = (|S(j)| - 1) + 2^{|S(j)|-1} - \sum_{\substack{i \in S(j) \\ i \neq n}} 2^{|S(i)|-1}.$$

Summing over all possible values of j , and adding 1 for the empty quotient, we get

the following bound on the complexity of A_S :

$$1 + \sum_{j=1}^{n-1} \left((|S(j)| - 1) + 2^{|S(j)|-1} - \sum_{\substack{i \in S(j) \\ i \neq n}} 2^{|S(i)|-1} \right).$$

Noting that $S(1) = \{2, \dots, n\}$ and $|S(1)| = n - 1$, we pull out the $j = 1$ case from the outermost summation:

$$1 + (n - 2) + 2^{n-2} - \sum_{\substack{i \in S(1) \\ i \neq n}} 2^{|S(i)|-1} + \sum_{j=2}^{n-1} \left((|S(j)| - 1) + 2^{|S(j)|-1} - \sum_{\substack{i \in S(j) \\ i \neq n}} 2^{|S(i)|-1} \right).$$

Observe that $1 + (n - 2) + 2^{n-2}$ is equal to $2^{n-2} + n - 1$, the bound we are trying to prove. We will show that the value of the rest of this formula is always less than or equal to zero. We pull $\sum_{j=2}^{n-1} 2^{|S(j)|-1}$ out to the front:

$$2^{n-2} + n - 1 + \sum_{j=2}^{n-1} 2^{|S(j)|-1} - \sum_{\substack{i \in S(1) \\ i \neq n}} 2^{|S(i)|-1} + \sum_{j=2}^{n-1} \left((|S(j)| - 1) - \sum_{\substack{i \in S(j) \\ i \neq n}} 2^{|S(i)|-1} \right).$$

Note that $\sum_{j=2}^{n-1} 2^{|S(j)|-1} = \sum_{\substack{i \in S(1) \\ i \neq n}} 2^{|S(i)|-1}$, so cancellation occurs:

$$2^{n-2} + n - 1 + \sum_{j=2}^{n-1} \left((|S(j)| - 1) - \sum_{\substack{i \in S(j) \\ i \neq n}} 2^{|S(i)|-1} \right).$$

Now, the value of the innermost summation is always greater than or equal to $|S(j)| - 1$: for each $i \in S(j)$, $i \neq n$, we know that n is a successor of i , and hence $S(i) \geq 1$ and $2^{|S(i)|-1} \geq 1$. Thus the value of the outermost summation is always less than or equal to zero. It follows that the number of quotients of A_S is at most $2^{n-2} + n - 1$. \square

Next we address the question of tightness of the bounds for two-sided ideals. For $n = 1$, $L = \Sigma^*$ is a two-sided ideal with one atom of complexity 1; so the bound of Proposition 5 does not hold.

The DFAs of Definition 5 and Figure 4 were introduced in [10]. It was shown in [7] that these languages have the largest syntactic semigroups amongst two-sided ideals of complexity n . Moreover, it was shown in [5] that they also meet the bounds on the quotient complexity of boolean operations, concatenation and star. Together with our result about the number of atoms and their complexity, this shows that these languages are “most complex” two-sided ideals.

Definition 5. Let $n \geq 4$, and let $\mathcal{D}_n = (Q_n, \Sigma, \delta_n, 1, \{n\})$ be the DFA with $\Sigma = \{a, b, c, d, e, f\}$, $a = (2, 3, \dots, n-1)$, $b = (2, 3)$, $c = (n-1 \rightarrow 2)$, $d = (n-1 \rightarrow 1)$, $e = (Q_{n-1} \rightarrow 2)$, and $f = (2 \rightarrow n)$. For $n = 4$, inputs a and b coincide. Also, let $\mathcal{D}_3 = (Q_3, \{a, b, c\}, \delta_3, 1, \{3\})$, where $a = 1$, $b = (Q_2 \rightarrow 2)$, $c = (2 \rightarrow 3)$, and let $\mathcal{D}_2 = (Q_2, \{a, b, c\}, \delta_2, 1, \{2\})$, where $a = 1$, $b = (Q_2 \rightarrow 2)$, $c = (Q_2 \rightarrow 1)$. Let L_n be the language accepted by \mathcal{D}_n .

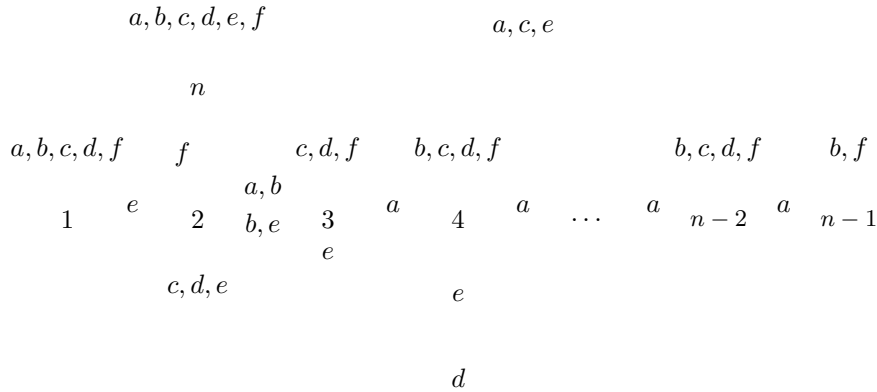


Figure 4: DFA of a two-sided ideal whose atoms meet the bounds.

Theorem 4. For $n \geq 2$, the language L_n of Definition 5 is a two-sided ideal that has $2^{n-2} + 1$ atoms and each atom meets the bounds of Proposition 5.

Proof. It was proved in [10] that L_n is a two-sided ideal of complexity n . The cases with $n < 4$ are easily verified; hence assume $n \geq 4$.

The following observations were made in [7]: Transformations $\{a, b, c\}$ restricted to $Q_n \setminus \{1, n\}$ generate all the transformations of $\{2, \dots, n-1\}$. Together with d and f , they generate all transformations of Q_n that fix 1 and n . Also, we have $ef = (Q_n \rightarrow n)$.

Recall that if A_S is an atom of a two-sided ideal, then $n \in S$, and either $S = Q_n$ or $1 \notin S$. We know A_{Q_n} is an atom of complexity n for all left ideals (and hence all two-sided ideals), so assume $n \in S$, $1 \notin S$. Then $1 \in \bar{S}$, and so from state (S, \bar{S}) in \mathcal{D}_S we can reach the final state $(\{n\}, \{1\})$ by transformations that fix 1 and n . Hence A_S is an atom for every S with $n \in S$, $1 \notin S$. There are 2^{n-2} of these atoms, as well as the atom A_{Q_n} , for a total of $2^{n-2} + 1$.

Consider the atom A_S for $S \neq Q_n$ and $S \neq Q_n \setminus \{1\}$. In the DFA \mathcal{D}_S , the initial state is (S, \bar{S}) , and we have $n \in S$, $1 \notin S$. By transformations that fix 1 and n , we can reach (X, Y) for all $X, Y \subseteq Q_n$ such that $n \in X$, $1 \in Y$, $X \cap Y = \emptyset$, $1 \leq |X| \leq |S|$, $1 \leq |Y| \leq n - |S|$. There are $\sum_{x=1}^{|S|} \sum_{y=1}^{n-|S|} \binom{n-2}{x-1} \binom{n-x-1}{y-1}$ such states. Since $n \in X$, $1 \notin X$ and $n \in \bar{Y}$, $1 \notin \bar{Y}$ we see that A_X and $A_{\bar{Y}}$ are atoms. Hence by the Distinguishability Lemma, all of these states are distinguishable from

each other and from \perp . Since $S \neq \emptyset$, we can reach \perp from (S, \overline{S}) by $ef = (Q_n \rightarrow n)$. Hence the bound is met.

It remains to show that the complexity of A_S , $S = Q_n \setminus \{1\}$ also meets the bound. The initial state of \mathcal{D}_S is $(\{2, \dots, n\}, \{1\})$. By transformations that fix 1 and n , we can reach all 2^{n-2} states of the form $(X, \{1\})$ with $\{n\} \subseteq X \subseteq Q_n \setminus \{1\}$. From $(\{n\}, \{1\})$, we can reach $n-2$ additional states $(\{n\}, \{i\})$ for $2 \leq i \leq n-1$ by ea^{i-2} . Finally, we can reach the sink state \perp from the initial state by $ef = (Q_n \rightarrow n)$. This gives a total of $2^{n-2} + n - 1$ reachable states, which matches the upper bound.

To see these states are distinguishable, note that A_X is an atom if $\{n\} \subseteq X \subseteq Q_n \setminus \{1\}$. Also, $A_{\overline{\{1\}}}$ is an atom. Hence by the Distinguishability Lemma, all states of the form $(X, \{1\})$ are distinguishable from each other and from \perp . Also, $(\{n\}, \{i\})$ is distinguished from $(\{n\}, \{j\})$ by $a^{n-i}f$, which sends the former state to the non-final state \perp , but sends the latter to some final state $(\{n\}, \{k\})$ with $k \neq 2$. And each $(\{n\}, \{j\})$, $1 \leq j \leq n-1$ is a final state, so it is distinguishable from all states of the form $(X, \{1\})$, $X \neq \{n\}$ and from \perp , since they are not final. Hence all $2^{n-2} + n - 1$ reachable states are distinguishable. \square

8 Some Numerical Results

The following tables compare the maximal complexities for atoms A_S of two-sided ideals (first entry), left ideals (second entry) and regular languages (third entry) with complexity n . Right ideals are omitted because their complexities are essentially the same as those of left ideals, by Remark 1. When the maximal complexity is undefined (e.g., because no languages in a class have atoms A_S for a particular size of S) this is indicated by an asterisk. The maximum values for each n are in boldface. The n^{th} entry in the *ratio* row shows the approximate value of m_n/m_{n-1} , where m_i is the i^{th} entry in the *max* row. It has been shown by Diekert and Walter [11] that the ratio converges exponentially fast to 3 for the class of regular languages and for all three classes of ideal languages.

n	1	2	3	4	5	...
$ S = 0$	$*/\mathbf{1/1}$	$*/\mathbf{2/3}$	$*/4/7$	$*/8/15$	$*/16/31$...
$ S = 1$	$\mathbf{1/1/1}$	$\mathbf{2/2/3}$	$3/5/\mathbf{10}$	$5/13/29$	$9/33/76$...
$ S = 2$		$\mathbf{2/2/3}$	$\mathbf{4/4/10}$	$\mathbf{8/16/43}$	$\mathbf{20/53/141}$...
$ S = 3$			$3/3/7$	$7/8/29$	$\mathbf{20/43/141}$...
$ S = 4$				$4/4/15$	$12/16/76$...
$ S = 5$					$5/5/31$...
<i>max</i>	$1/1/1$	$2/2/3$	$4/5/10$	$8/16/43$	$20/53/141$...
<i>ratio</i>	–	$2.00/2.00/3.00$	$2.00/2.50/3.33$	$2.00/3.20/4.30$	$2.50/3.31/3.28$...

n	6	7	8	9
$ S = 0$	*/32/63	*/64/127	*/128/255	*/256/511
$ S = 1$	17/81/187	33/193/442	65/449/1,017	129/1,025/2,296
$ S = 2$	48/156/406	112/427/1,086	256/1,114/2,773	576/2,809/6,859
$ S = 3$	64/166/501	182/542/1,548	484/1,611/4,425	1,234/4,517/12,043
$ S = 4$	48/106/406	182/462/1,548	584/1,646/5,083	1,710/5,245/15,361
$ S = 5$	21/32/187	112/249/1,086	484/1,205/4,425	1,710/4,643/15,361
$ S = 6$	6/6/63	38/64/442	256/568/2,773	1,234/3,019/12,043
$ S = 7$		7/7/127	71/128/1,017	576/1,271/6,859
$ S = 8$			8/8/255	136/256/2,296
$ S = 9$				9/9/511
<i>max</i>	64/166/501	182/542/1,548	584/1,646/5,083	1,710/5,245/15,361
<i>ratio</i>	3.20/3.13/3.55	2.84/3.27/3.09	3.21/3.04/3.28	2.93/3.19/3.02

9 Conclusions

We have derived tight upper bounds for the number of atoms and quotient complexity of atoms in right, left and two-sided regular ideal languages. The recently discovered relationship between atoms and the Myhill and Nerode congruence classes opens up many interesting research questions. The quotient complexity of a language is equal to the number of Nerode classes, and the number of Myhill classes has also been used as a measure of complexity, called *syntactic complexity* since it is equal to the size of the syntactic semigroup. We can view the number of atoms as a third fundamental measure of complexity for regular languages.

It is known [8] that the number of atoms of a regular language L is equal to the quotient complexity of the *reversal* of L . The quotient complexity of reversal has been studied for various classes of languages in the context of determining the quotient complexity of operations on regular languages. Hence, the maximal number of atoms is known for many language classes.

However, as far as we know the *quotient complexity* of atoms has not been studied outside of regular languages and ideals. For simplicity, let us call the atom congruence the *left congruence*, the Nerode congruence the *right congruence*, and the Myhill congruence the *central congruence*. When computing the quotient complexity of atoms, we are computing the number of *right congruence classes* of each *left congruence class*. We can consider variations of this idea: how many right classes and left classes do the central classes have? How many central classes do the left classes have? These questions are outside the scope of this paper, but we believe they should be investigated.

References

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

- [2] Brzozowski, J. In search of the most complex regular languages. *Int. J. Found. Comput. Sci.*, 24(6):691–708, 2013.
- [3] Brzozowski, J. and Davies, G. Maximally atomic languages. In Ésik, Z. and Fülöp, Z., editors, *Proceedings of the 14th International Conference on Automata and Formal Languages (AFL)*, volume 151 of *Electronic Proceedings in Theoretical Computer Science*, pages 151–161, 2014.
- [4] Brzozowski, J. and Davies, G. Most complex regular right ideals. In Jürgensen, H. and et al., editors, *Proceedings of the 16th International Workshop on Descriptive Complexity of Formal Systems (DCFS)*, volume 8614 of *Lecture Notes in Computer Science*, pages 90–101, Berlin/Heidelberg, 2014. Springer.
- [5] Brzozowski, J., Davies, S., and Liu, B. Y. V. Most complex regular ideal languages. <http://arxiv.org/abs/1511.00157>, November, 2015.
- [6] Brzozowski, J., Jirásková, G., and Li, B. Quotient complexity of ideal languages. *Theoret. Comput. Sci.*, 470:36–52, 2013.
- [7] Brzozowski, J. and Szykuła, M. Upper bounds on syntactic complexity of left and two-sided ideals. In Shur, A. M. and Volkov, M. V., editors, *Proceedings of the 18th International Conference on Developments in Language Theory (DLT)*, volume 8633 of *Lecture Notes in Computer Science*, pages 13–24, Berlin/Heidelberg, 2014. Springer.
- [8] Brzozowski, J. and Tamm, H. Quotient complexities of atoms of regular languages. *Int. J. Found. Comput. Sci.*, 24(7):1009–1027, 2013.
- [9] Brzozowski, J. and Tamm, H. Theory of atomata. *Theoret. Comput. Sci.*, 539:13–27, 2014.
- [10] Brzozowski, J. and Ye, Y. Syntactic complexity of ideal and closed languages. In Mauri, G. and Leporati, A., editors, *Proceedings of the 15th International Conference on Developments in Language Theory (DLT)*, volume 6795 of *Lecture Notes in Computer Science*, pages 117–128, Berlin/Heidelberg, 2011. Springer.
- [11] Diekert, V. and Walter, T. Asymptotic approximation for the quotient complexities of atoms. *Acta Cybernetica*, 22(2):349–357, 2015.
- [12] Iván, S. Complexity of atoms, combinatorially. <http://arxiv.org/abs/1404.6632>, 2015.
- [13] Myhill, J. Finite automata and representation of events. *Wright Air Development Center Technical Report*, 57–624, 1957.
- [14] Nerode, A. Linear automaton transformations. *Proc. Amer. Math. Soc.*, 9:541–544, 1958.

- [15] Pin, Jean-Eric. Syntactic semigroups. In *Handbook of Formal Languages, vol. 1: Word, Language, Grammar*, pages 679–746. Springer, New York, NY, USA, 1997.
- [16] Yu, Sheng. State complexity of regular languages. *J. Autom. Lang. Comb.*, 6:221–234, 2001.

Received 15th June 2015