Quotient Complexities of Atoms of Regular Languages*

Janusz Brzozowski[1] and Hellis Tamm[2]

[1] David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, ON, Canada N2L 3G1
brzozo@uwaterloo.ca

[2] Institute of Cybernetics, Tallinn University of Technology, Akadeemia tee 21, 12618 Tallinn, Estonia
hellis@cs.ioc.ee

Abstract. An atom of a regular language L with n (left) quotients is a non-empty intersection of uncomplemented or complemented quotients of L, where each of the n quotients appears in a term of the intersection. The quotient complexity of L, which is the same as the state complexity of L, is the number of quotients of L. We prove that, for any language L with quotient complexity n, the quotient complexity of any atom of L with r complemented quotients has an upper bound of $2^n - 1$ if $r = 0$ or $r = n$, and $1 + \sum_{k=1}^{r} \sum_{h=k+1}^{k+n-r} C_h^n \cdot C_k^h$ otherwise, where C_j^i is the binomial coefficient. For each $n \geqslant 1$, we exhibit a language whose atoms meet these bounds.

1 Introduction

Atoms of regular languages were introduced in 2011 by Brzozowski and Tamm [3]; we briefly state their main properties here.

If Σ is a non-empty finite alphabet, then Σ^* is the free monoid generated by Σ. A *word* is any element of Σ^*, and the empty word is ε. A *language* over Σ is any subset of Σ^*. The *reverse of a language* L is denoted by L^R and defined as $L^R = \{w^R \mid w \in L\}$, where w^R is w spelled backwards.

The *(left) quotient* of a regular language L over an alphabet Σ by a word $w \in \Sigma^*$ is the language $w^{-1}L = \{x \in \Sigma^* \mid wx \in L\}$. It is well known that a language L is regular if and only if it has a finite number of distinct quotients, and that the number of states in the minimal deterministic finite automaton (DFA) recognizing L is precisely the number of distinct quotients of L. Also, L is its own quotient by the empty word ε, that is $\varepsilon^{-1}L = L$. Note too that the quotient by $u \in \Sigma^*$ of the quotient by $w \in \Sigma^*$ of L is the quotient by wu of L, that is, $u^{-1}(w^{-1}L) = (wu)^{-1}L$.

* This work was supported by the Natural Sciences and Engineering Research Council of Canada under grant No. OGP0000871, the ERDF funded Estonian Center of Excellence in Computer Science, EXCS, the Estonian Science Foundation grant 7520, and the Estonian Ministry of Education and Research target-financed research theme no. 0140007s12.

H.-C. Yen and O.H. Ibarra (Eds.): DLT 2012, LNCS 7410, pp. 50–61, 2012.

An *atom*[1] of a regular language L with quotients $K_0, \ldots, K_{n-1}$ is any non-empty language of the form $\widetilde{K_0} \cap \cdots \cap \widetilde{K_{n-1}}$, where $\widetilde{K_i}$ is either K_i or $\overline{K_i}$, and $\overline{K_i}$ is the complement of K_i with respect to Σ^*. Thus atoms of L are regular languages uniquely determined by L and they define a partition of Σ^*. They are pairwise disjoint, every quotient of L (including L itself) is a union of atoms, and every quotient of an atom is a union of atoms. Thus the atoms of a regular language are its basic building blocks. Also, $\overline{L}$ defines the same atoms as L.

The *quotient complexity* [2] of L is the number of quotients of L, and this is the same number as the number of states in the minimal DFA recognizing L; the latter number is known as the *state complexity* [7] of L. Quotient complexity allows us to use language-theoretic methods, whereas state complexity is more amenable to automaton-theoretic techniques. We use one of these two points of view or the other, depending on convenience.

We study the quotient complexity of atoms of regular languages. Our main result is the following:

Theorem 1 (Main Result). *Let $L \subseteq \Sigma^*$ be a non-empty regular language and let its set of quotients be $K = \{K_0, K_1, \ldots, K_{n-1}\}$. For $n \geqslant 1$, the quotient complexity of the atoms with 0 or n complemented quotients is less than or equal to $2^n - 1$. For $n \geqslant 2$ and r satisfying $1 \leqslant r \leqslant n-1$, the quotient complexity of any atom of L with r complemented quotients is less than or equal to*

$$f(n,r) = 1 + \sum_{k=1}^{r} \sum_{h=k+1}^{k+n-r} C_h^n \cdot C_k^h,$$

where C_j^i is the binomial coefficient "i choose j". For $n = 1$, the single atom Σ^* of the language Σ^* or $\emptyset$ meets the bound 1. Moreover, for $n \geqslant 2$, all the atoms of the language L_n recognized by the DFA $\mathcal{D}_n$ of Figure 1 meet these bounds.

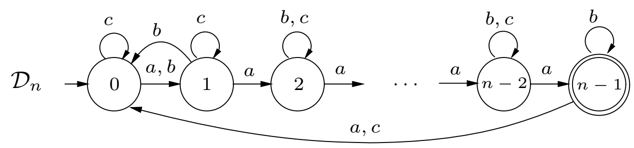

Fig. 1. DFA $\mathcal{D}_n$ of language L_n whose atoms meet the bounds

In Section 2 we derive upper bounds on the quotient complexities of atoms. In Section 3 we define our notation and terminology for automata, and present the definition of the átomaton [3] of a regular language; this is a nondeterministic finite automaton (NFA) whose states are the atoms of the language. We also provide a different characterization of the átomaton. We introduce a class of DFA's in Section 4 and study the átomata of their languages. We then prove

[1] The definition in [3] does not consider the intersection of all the complemented quotients to be an atom. Our new definition adds symmetry to the theory.

in Section 5 that the atoms of these languages meet the quotient complexity bounds. Section 6 concludes the paper. Proofs that are omitted can be found at `http://arxiv.org/abs/1201.0295`.

2 Upper Bounds on the Quotient Complexities of Atoms

We first derive upper bounds on the quotient complexity of atoms. We use quotients here, since they are convenient for this task. First we deal with the two atoms that have only uncomplemented or only complemented quotients.

Let $L \subseteq \Sigma^*$ be a non-empty regular language and let its set of quotients be $K = \{K_0, K_1, \ldots, K_{n-1}\}$, with $n \geqslant 1$.

Proposition 1 (Atoms with 0 or n Complemented Quotients) *For $n \geqslant 1$, the quotient complexity of the two atoms $A_K = K_0 \cap \cdots \cap K_{n-1}$ and $A_\emptyset = \overline{K_0} \cap \cdots \cap \overline{K_{n-1}}$ is less than or equal to $2^n - 1$.*

Proof. Every quotient $w^{-1}A_K$ of atom A_K is the intersection of languages $w^{-1}K_i$, which are quotients of L:

$$w^{-1}A_K = w^{-1}(K_0 \cap \cdots \cap K_{n-1}) = w^{-1}K_0 \cap \cdots \cap w^{-1}K_{n-1}.$$

Since these quotients of L need not be distinct, $w^{-1}A_K$ may be the intersection of any non-empty subset of quotients of L. Hence A_K can have at most $2^n - 1$ quotients.

The argument for the atom $A_\emptyset = \overline{K_0} \cap \cdots \cap \overline{K_{n-1}}$ with n complemented quotients is similar, since $w^{-1}\overline{K_i} = \overline{w^{-1}K_i}$. □

Next, we present an upper bound on the quotient complexity of any atom with at least one and fewer than n complemented quotients.

Proposition 2 (Atoms with r Complemented Quotients, $1 \leqslant r \leqslant n-1$). *For $n \geqslant 2$ and $1 \leqslant r \leqslant n-1$, the quotient complexity of any atom with r complemented quotients is less than or equal to*

$$f(n,r) = 1 + \sum_{k=1}^{r} \sum_{h=k+1}^{k+n-r} C_h^n \cdot C_k^h. \tag{1}$$

Proof. Consider an intersection of complemented and uncomplemented quotients that constitutes an atom. Without loss of generality, we arrange the terms in the intersection in such a way that all complemented quotients appear on the right. Thus let $A_i = K_0 \cap \cdots \cap K_{n-r-1} \cap \overline{K_{n-r}} \cap \cdots \cap \overline{K_{n-1}}$ be an atom of L with r complemented quotients of L, where $1 \leqslant r \leqslant n-1$. The quotient of A_i by any word $w \in \Sigma^*$ is

$$\begin{aligned} w^{-1}A_i &= w^{-1}(K_0 \cap \cdots \cap K_{n-r-1} \cap \overline{K_{n-r}} \cap \cdots \cap \overline{K_{n-1}}) \\ &= w^{-1}K_0 \cap \cdots \cap w^{-1}K_{n-r-1} \cap \overline{w^{-1}K_{n-r}} \cap \cdots \cap \overline{w^{-1}K_{n-1}}. \end{aligned}$$

Since each quotient $w^{-1}K_j$ is a quotient, say K_{i_j}, of L, we have

$$w^{-1}A_i = K_{i_0} \cap \cdots \cap K_{i_{n-r-1}} \cap \overline{K_{i_{n-r}}} \cap \cdots \cap \overline{K_{i_{n-1}}}.$$

The cardinality of a set S is denoted by $|S|$. Let the set of distinct quotients of L appearing in $w^{-1}A_i$ uncomplemented (respectively, complemented) be X (respectively, Y), where $1 \leqslant |X| \leqslant n - r$ and $1 \leqslant |Y| \leqslant r$. If $X \cap Y \neq \emptyset$, then $w^{-1}A_i = \emptyset$. Therefore assume that $X \cap Y = \emptyset$, and that $|X \cup Y| = h$, where $2 \leqslant h \leqslant n$; there are C^n_h such sets $X \cup Y$. Suppose further that $|Y| = k$, where $1 \leqslant k \leqslant r$. There are C^h_k ways of choosing Y. Hence there are at most $\sum_{h=k+1}^{k+n-r} C^n_h \cdot C^h_k$ distinct intersections with k complemented quotients. Thus, the total number of intersections of uncomplemented and complemented quotients can be at most $\sum_{k=1}^{r} \sum_{h=k+1}^{k+n-r} C^n_h \cdot C^h_k$.

Adding 1 for the empty quotient of $w^{-1}A_i$, we get the required bound. □

We now consider the properties of the function $f(n, r)$.

Proposition 3 (Properties of Bounds). *For any $n \geqslant 2$ and $1 \leqslant r \leqslant n-1$, the function $f(n, r)$ of Equation (1) satisfies the following properties:*

1. *$f(n, r) = f(n, n - r)$.*
2. *For a fixed n, the maximal value of $f(n, r)$ occurs when $r = \lfloor n/2 \rfloor$.*

Some numerical values of $f(n, r)$ are shown in Table 1. The figures in boldface type are the maxima for a fixed n. The row marked *max* shows the maximal quotient complexity of the atoms of L. The row marked *ratio* shows the value of $f(n, \lfloor n/2 \rfloor)/f(n-1, \lfloor (n-1)/2 \rfloor)$, for $n \geqslant 2$. It appears that this ratio converges to 3. For example, for $n = 100$ it is approximately 3.0002.

Table 1. Maximal quotient complexity of atoms

n	1	2	3	4	5	6	7	8	9	10	···
r=0	**1**	**3**	7	15	31	63	127	255	511	1,023	···
r=1	**1**	**3**	**10**	29	76	187	442	1,017	2,296	5,111	···
r=2	*	**3**	**10**	**43**	**141**	406	1,086	2,773	6,859	16,576	···
r=3	*	*	7	29	**141**	**501**	**1,548**	4,425	12,043	31,681	···
r=4	*	*	*	15	76	406	**1,548**	**5,083**	**15,361**	44,071	···
r=5	*	*	*	*	31	187	1,086	4,425	**15,361**	**48,733**	···
max	1	3	10	43	141	501	1,548	5,083	15,361	48,733	···
ratio	–	3	3.33	4.30	3.28	3.55	3.09	3.28	3.02	3.17	···

3 Automata and Átomata of Regular Languages

A *nondeterministic finite automaton (NFA)* is a quintuple $\mathcal{N} = (Q, \Sigma, \eta, I, F)$, where Q is a finite, non-empty set of *states*, Σ is a finite non-empty *alphabet*, $\eta : Q \times \Sigma \to 2^Q$ is the *transition function*, $I \subseteq Q$ is the set of *initial states*, and

$F \subseteq Q$ is the set of *final states*. As usual, we extend the transition function to functions $\eta' : Q \times \Sigma^* \to 2^Q$, and $\eta'' : 2^Q \times \Sigma^* \to 2^Q$. We do not distinguish these functions notationally, but use η for all three.

The *language accepted* by an NFA $\mathcal{N}$ is $L(\mathcal{N}) = \{w \in \Sigma^* \mid \eta(I, w) \cap F \neq \emptyset\}$. Two NFA's are *equivalent* if they accept the same language. The *right language* of a state q of $\mathcal{N}$ is $L_{q,F}(\mathcal{N}) = \{w \in \Sigma^* \mid \eta(q, w) \cap F \neq \emptyset\}$. The *right language* of a set S of states of $\mathcal{N}$ is $L_{S,F}(\mathcal{N}) = \bigcup_{q \in S} L_{q,F}(\mathcal{N})$; hence $L(\mathcal{N}) = L_{I,F}(\mathcal{N})$. A state is *empty* if its right language is empty. Two states of an NFA are *equivalent* if their right languages are equal. The *left language* of a state q of $\mathcal{N}$ is $L_{I,q} = \{w \in \Sigma^* \mid q \in \eta(I, w)\}$. A state is *unreachable* if its left language is empty. An NFA is *trim* if it has no empty or unreachable states.

A *deterministic finite automaton (DFA)* is a quintuple $\mathcal{D} = (Q, \Sigma, \delta, q_0, F)$, where Q, Σ, and F are as in an NFA, $\delta : Q \times \Sigma \to Q$ is the transition function, and q_0 is the initial state. We use the following operations on automata:

1. The *determinization* operation $\mathbb{D}$ applied to an NFA $\mathcal{N}$ yields a DFA $\mathcal{N}^{\mathbb{D}}$ obtained by the subset construction, where only subsets reachable from the initial subset of $\mathcal{N}^{\mathbb{D}}$ are used and the empty subset, if present, is included.
2. The *reversal* operation $\mathbb{R}$ applied to an NFA $\mathcal{N}$ yields an NFA $\mathcal{N}^{\mathbb{R}}$, where sets of initial and final states of $\mathcal{N}$ are interchanged and each transition between any two states is reversed.

Let L be any non-empty regular language, and let its set of quotients be $K = \{K_0, \ldots, K_{n-1}\}$. One of the quotients of L is L itself; this is called the *initial* quotient and is denoted by K_{in}. A quotient is *final* if it contains the empty word ε. The set of final quotients is $F = \{K_i \mid \varepsilon \in K_i\}$.

In the following definition we use a one-to-one correspondence $K_i \leftrightarrow \mathbf{K}_i$ between quotients K_i of a language L and the states $\mathbf{K}_i$ of the *quotient DFA* $\mathcal{D}$ defined below. We refer to the $\mathbf{K}_i$ as *quotient symbols*.

Definition 1. *The* quotient DFA *of* L *is* $\mathcal{D} = (\mathbf{K}, \Sigma, \delta, \mathbf{K}_{in}, \mathbf{F})$*, where* $\mathbf{K} = \{\mathbf{K}_0, \ldots, \mathbf{K}_{n-1}\}$*,* $\mathbf{K}_{in}$ *corresponds to* K_{in}*,* $\mathbf{F} = \{\mathbf{K}_i \mid K_i \in F\}$*, and* $\delta(\mathbf{K}_i, a) = \mathbf{K}_j$ *if and only if* $a^{-1}K_i = K_j$*, for all* $\mathbf{K}_i, \mathbf{K}_j \in \mathbf{K}$ *and* $a \in \Sigma$*.*

In a quotient DFA the right language of $\mathbf{K}_i$ is K_i, and its left language is $\{w \in \Sigma^* \mid w^{-1}L = K_i\}$. The language $L(\mathcal{D})$ is the right language of $\mathbf{K}_{in}$, and hence $L(\mathcal{D}) = L$. Also, DFA $\mathcal{D}$ is minimal, since all quotients in K are distinct.

It follows from the definition of an atom, that a regular language L has at most 2^n atoms. An atom is *initial* if it has L (rather than $\overline{L}$) as a term; it is *final* if it contains ε. Since L is non-empty, it has at least one quotient containing ε. Hence it has exactly one final atom, the atom $\widehat{K_0} \cap \cdots \cap \widehat{K_{n-1}}$, where $\widehat{K_i} = K_i$ if $\varepsilon \in K_i$, and $\widehat{K_i} = \overline{K_i}$ otherwise. Let $A = \{A_0, \ldots, A_{m-1}\}$ be the set of atoms of L. By convention, I is the set of initial atoms and A_{m-1} is the final atom.

As above, we use a one-to-one correspondence $A_i \leftrightarrow \mathbf{A}_i$ between atoms A_i of a language L and the states $\mathbf{A}_i$ of the NFA $\mathcal{A}$ defined below. We refer to the $\mathbf{A}_i$ as *atom symbols*.

Definition 2. *The* átomaton[2] *of L is the NFA $\mathcal{A} = (\mathbf{A}, \Sigma, \eta, \mathbf{I}, \{\mathbf{A}_{m-1}\})$, where $\mathbf{A} = \{\mathbf{A}_i \mid A_i \in A\}$, $\mathbf{I} = \{\mathbf{A}_i \mid A_i \in I\}$, $\mathbf{A}_{m-1}$ corresponds to A_{m-1}, and $\mathbf{A}_j \in \eta(\mathbf{A}_i, a)$ if and only if $aA_j \subseteq A_i$, for all $\mathbf{A}_i, \mathbf{A}_j \in \mathbf{A}$ and $a \in \Sigma$.*

Example 1. Let $L_2 \subseteq \{a, c\}^*$ be defined by the quotient equations below (left) and recognized by the DFA $\mathcal{D}_2$ of Fig. 2 (a). The equations for the atoms of L_2 are below (right), and the átomaton $\mathcal{A}_2$ is in Fig. 2 (b); here each atom is denoted by A_P, where P is the set of uncomplemented quotients. Thus $K_0 \cap \overline{K_1}$ becomes $A_{\{0\}}$, etc., and we represent the sets in the subscripts without brackets and commas. The reverse $\mathcal{D}_2^{\mathbb{R}}$ of $\mathcal{D}_2$ is in Fig. 2 (c). The determinized reverse $\mathcal{D}_2^{\mathbb{RD}}$ is in Fig. 2 (d); this is the minimal DFA for L_2^R, the reverse of L_2. The reverse $\mathcal{A}_2^{\mathbb{R}}$ of the átomaton is in Fig. 2 (e). Note that $\mathcal{D}_2^{\mathbb{RD}}$ and $\mathcal{A}_2^{\mathbb{R}}$ are isomorphic.

$$K_0 = aK_1 \cup cK_0, \qquad K_0 \cap K_1 = a(K_0 \cap K_1) \cup c[(K_0 \cap K_1) \cup (K_0 \cap \overline{K_1})],$$
$$K_1 = aK_0 \cup cK_0 \cup \varepsilon, \qquad K_0 \cap \overline{K_1} = a(\overline{K_0} \cap K_1),$$
$$\overline{K_0} \cap K_1 = a(K_0 \cap \overline{K_1}) \cup \varepsilon,$$
$$\overline{K_0} \cap \overline{K_1} = a(\overline{K_0} \cap \overline{K_1}) \cup c[(\overline{K_0} \cap \overline{K_1}) \cup (\overline{K_0} \cap K_1)].$$

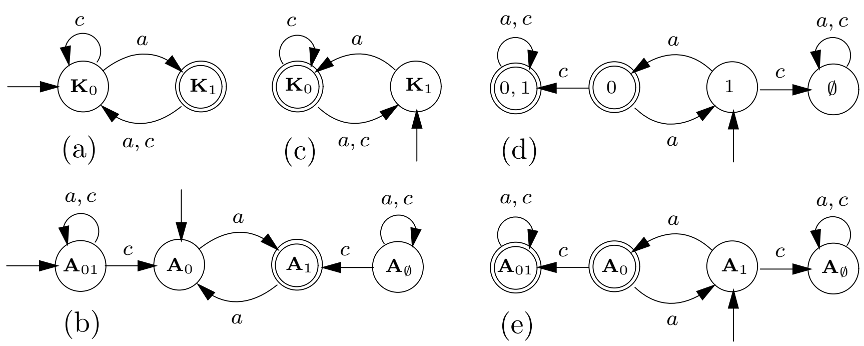

Fig. 2. (a) DFA $\mathcal{D}_2$; (b) Átomaton $\mathcal{A}_2$; (c) NFA $\mathcal{D}_2^{\mathbb{R}}$; (d) DFA $\mathcal{D}_2^{\mathbb{RD}}$; (e) DFA $\mathcal{A}_2^{\mathbb{R}}$

The next theorem from [1], also discussed in [3], will be used several times.

Theorem 2 (Determinization). *If an NFA $\mathcal{N}$ has no empty states and $\mathcal{N}^{\mathbb{R}}$ is deterministic, then $\mathcal{N}^{\mathbb{D}}$ is minimal.*

It was shown in [3] that the átomaton $\mathcal{A}$ of L with reachable atoms only is isomorphic to the trimmed version of $\mathcal{D}^{\mathbb{RDR}}$, where $\mathcal{D}$ is the quotient DFA of L.

[2] In [3], the intersection $A_\emptyset = \overline{K_0} \cap \cdots \cap \overline{K_{n-1}}$ was not considered an atom. It was shown that the right language of state $\mathbf{A}_i$ is the atom A_i, the left language of $\mathbf{A}_i$ is non-empty, the language of the átomaton $\mathcal{A}$ is L, and $\mathcal{A}$ is trim. If the intersection $A_\emptyset$ of all the complemented quotients is non-empty, then $A_\emptyset$ is an atom and $\mathcal{A}$ is no longer trim because state $\mathbf{A}_\emptyset$ is not reachable from any initial state.

With our new definition, $\mathcal{A}$ is isomorphic to $\mathcal{D}^{\mathbb{RDR}}$. We now study this isomorphism in detail, along with the isomorphism between $\mathcal{A}^{\mathbb{R}}$ and $\mathcal{D}^{\mathbb{RD}}$. We deal with the following automata:

1. Quotient DFA $\mathcal{D} = (\mathbf{K}, \Sigma, \delta, \mathbf{K}_{in}, \mathbf{F})$ of L whose states are *quotient symbols*.
2. The reverse $\mathcal{D}^{\mathbb{R}} = (\mathbf{K}, \Sigma, \delta^{\mathbb{R}}, \mathbf{F}, \{\mathbf{K}_{in}\})$ of $\mathcal{D}$. The states in $\mathbf{K}$ are still *quotient symbols,* but their right languages are no longer quotients of L.
3. The determinized reverse $\mathcal{D}^{\mathbb{RD}} = (S, \Sigma, \alpha, \mathbf{F}, G)$, where $S \subseteq 2^{\mathbf{K}}$ and $G = \{S_i \in S \mid \mathbf{K}_{in} \in S_i\}$. The states in S are *sets of quotient symbols*, i.e., subsets of $\mathbf{K}$. Since $(\mathcal{D}^{\mathbb{R}})^{\mathbb{R}} = \mathcal{D}$ is deterministic and all of its states are reachable, $\mathcal{D}^{\mathbb{R}}$ has no empty states. By Theorem 2, DFA $\mathcal{D}^{\mathbb{RD}}$ is minimal and accepts L^R; hence it is isomorphic to the quotient DFA of L^R.
4. The reverse $\mathcal{D}^{\mathbb{RDR}} = (S, \Sigma, \alpha^{\mathbb{R}}, G, \{\mathbf{F}\})$ of $\mathcal{D}^{\mathbb{RD}}$; here the states are still *sets of quotient symbols.*
5. The átomaton $\mathcal{A} = (\mathbf{A}, \Sigma, \eta, \mathbf{I}, \{\mathbf{A}_{m-1}\})$, whose states are *atom symbols.*
6. The reverse $\mathcal{A}^{\mathbb{R}} = (\mathbf{A}, \Sigma, \eta^{\mathbb{R}}, \mathbf{A}_{m-1}, \mathbf{I})$ of $\mathcal{A}$, whose states are still *atom symbols,* though their right languages are no longer atoms.

The results from [3] and our new definition of atoms imply that $\mathcal{A}^{\mathbb{R}}$ is a minimal DFA that accepts L^R. It follows that $\mathcal{A}^{\mathbb{R}}$ is isomorphic to $\mathcal{D}^{\mathbb{RD}}$. Our next result makes this isomorphism precise.

Proposition 4 (Isomorphism). *Let $\varphi : \mathbf{A} \to S$ be the mapping assigning to state $\mathbf{A}_j$, given by $A_j = K_{i_0} \cap \cdots \cap K_{i_{n-r-1}} \cap \overline{K_{i_{n-r}}} \cap \cdots \cap \overline{K_{i_{n-1}}}$ of $\mathcal{A}^{\mathbb{R}}$, the set $\{K_{i_0}, \ldots, K_{i_{n-r-1}}\}$. Then φ is a DFA isomorphism between $\mathcal{A}^{\mathbb{R}}$ and $\mathcal{D}^{\mathbb{RD}}$.*

Proof. The initial state $\mathbf{A}_{m-1}$ of $\mathcal{A}^{\mathbb{R}}$ is mapped to the set of all quotients containing ε, which is precisely the initial state $\mathbf{F}$ of $\mathcal{D}^{\mathbb{RD}}$. Since the quotient L appears uncomplemented in every initial atom $A_i \in I$, the image $\varphi(\mathbf{A}_i)$ contains L. Thus the set of final states of $\mathcal{A}^{\mathbb{R}}$ is mapped to the set of final states of $\mathcal{D}^{\mathbb{RD}}$.

It remains to be shown, for all $\mathbf{A}_i, \mathbf{A}_j \in \mathbf{A}$ and $a \in \Sigma$, that $\eta^{\mathbb{R}}(\mathbf{A}_j, a) = \mathbf{A}_i$ if and only if $\alpha(\varphi(\mathbf{A}_j), a) = \varphi(\mathbf{A}_i)$.

Consider atom A_i with P_i as the set of quotients that appear uncomplemented in A_i. Also define the corresponding set P_j for A_j. If there is a missing quotient K_h in the intersection $a^{-1}A_i$, we use $a^{-1}A_i \cap (K_h \cup \overline{K_h})$. We do this for all missing quotients until we obtain a union of atoms. Hence $\mathbf{A}_j \in \eta(\mathbf{A}_i, a)$ can hold in $\mathcal{A}$ if and only if $P_j \supseteq \delta(P_i, a)$ and $P_j \cap \delta(Q \setminus P_i, a) = \emptyset$. It follows that in $\mathcal{A}^{\mathbb{R}}$ we have $\eta^{\mathbb{R}}(\mathbf{A}_j, a) = \mathbf{A}_i$ if and only if $P_j \supseteq \delta(P_i, a)$ and $P_j \cap \delta(Q \setminus P_i, a) = \emptyset$.

Now consider $\mathcal{D}^{\mathbb{RD}}$. Let P_i be any subset of Q; then the successor set of P_i in $\mathcal{D}$ is $\delta(P_i, a)$. Let $\delta(P_i, a) = P_k$. So in $\mathcal{D}^{\mathbb{R}}$, we have $P_i \in \delta^{\mathbb{R}}(P_k, a)$. But suppose that state q is not in $\delta(Q, a)$; then $\delta^{\mathbb{R}}(q, a) = \emptyset$. Consequently, we also have $P_i \in \delta^{\mathbb{R}}(P_k \cup \{q\}, a)$. It follows that for any P_j containing $\delta(P_i, a)$ and satisfying $P_j \cap \delta(Q \setminus P_i, a) = \emptyset$, we also have $\alpha(P_j, a) = P_i$.

We have now shown that $\eta^{\mathbb{R}}(\mathbf{A}_j, a) = \mathbf{A}_i$ if and only if $\alpha(P_j, a) = P_i$, for all subsets $P_i, P_j \in S$, that is, if and only if $\alpha(\varphi(\mathbf{A}_j), a) = \varphi(\mathbf{A}_i)$. □

Corollary 1. *The mapping φ is an NFA isomorphism between $\mathcal{A}$ and $\mathcal{D}^{\mathbb{RDR}}$.*

In the remainder of the paper it is more convenient to use the $\mathcal{D}^{\mathbb{RDR}}$ representation of átomata, rather than that of Definition 2.

4 The Witness Languages and Automata

We now introduce a class $\{L_n \mid n \geqslant 2\}$ of regular languages defined by the quotient DFA's $\mathcal{D}_n$ given below; we shall prove that the atoms of each language $L_n = L(\mathcal{D}_n)$ in this class meet the worst-case quotient complexity bounds.

Definition 3 (Witness). *For $n \geqslant 2$, let $\mathcal{D}_n = (Q, \Sigma, \delta, q_0, F)$, where $Q = \{0, \dots, n-1\}$, $\Sigma = \{a, b, c\}$, $q_0 = 0$, $F = \{n-1\}$, $\delta(i, a) = i + 1 \bmod n$, $\delta(0, b) = 1$, $\delta(1, b) = 0$, $\delta(i, b) = i$ for $i > 1$, $\delta(i, c) = i$ for $0 \leqslant i \leqslant n-2$, and $\delta(n-1, c) = 0$. Let L_n be the language accepted by $\mathcal{D}_n$.*

For $n \geqslant 3$, the DFA of Definition 3 is illustrated in Fig. 1, and $\mathcal{D}_2$ is the DFA of Example 1 (a and b coincide). The DFA $\mathcal{D}_n$ is minimal, since for $0 \leqslant i \leqslant n-1$, state i accepts a^{n-1-i}, and no other state accepts this word.

A *transformation* of a set Q is a mapping of Q into itself. If t is a transformation of Q and $i \in Q$, then it is the *image* of i under t. The set of all transformations of a finite set Q is a semigroup under composition, in fact, a monoid $\mathcal{T}_Q$ of n^n elements. A *permutation* of Q is a mapping of Q *onto* itself. A *transposition* (i, j) interchanges i and j and does not affect any other elements. A *singular* transformation, denoted by $\binom{i}{j}$, has $it = j$ and $ht = h$ for all $h \neq i$.

In 1935 Piccard [5] proved that three transformations of Q are sufficient to generate $\mathcal{T}_Q$. Dénes [4] studied more general generators; we use his formulation:

Theorem 3 (Transformations). *The transformation monoid $\mathcal{T}_Q$ can be generated by any cyclic permutation of n elements together with any transposition and any singular transformation.*

In any DFA $\mathcal{D} = (Q, \Sigma, \delta, q_0, F)$, each word w in Σ^+ performs a transformation on Q defined by $\delta(\cdot, w)$. The set of all these transformations is the *transformation semigroup* of $\mathcal{D}$. By Theorem 3, the transformation semigroup of our witness $\mathcal{D}_n$ has n^n elements, since a is a cyclic permutation, b is a transposition and c is a singular transformation.

The following result of Salomaa, Wood and Yu [6] concerning reversal is restated in our terminology.

Theorem 4 (Transformations and Reversal). *Let $\mathcal{D}$ be a minimal DFA with n states accepting a language L. If the transformation semigroup of $\mathcal{D}$ has n^n elements, then the quotient complexity of L^R is 2^n.*

Corollary 2 (Reversal). *For $n \geqslant 2$, the quotient complexity of L_n^R is 2^n.*

Corollary 3 (Number of Atoms of L_n). *The language L_n has 2^n atoms.*

Proof. By Corollary 1, the átomaton of L_n is isomorphic to the reversed quotient DFA of L_n^R. By Corollary 2, the quotient DFA of L_n^R has 2^n states, and so the empty set of states of L_n is reachable in L_n^R. Hence L_n^R has the empty quotient, implying that the intersection of all the complemented quotients of L_n is non-empty, and so L_n has 2^n atoms. □

Proposition 5 (Transitions of the Átomaton). *Let $\mathcal{D}_n = (Q, \Sigma, \delta, q_0, F)$ be the DFA of Definition 3. The átomaton of $L_n = L(\mathcal{D}_n)$ is the NFA $\mathcal{A}_n = (2^Q, \Sigma, \eta, I, \{n-1\})$, where*

1. *If $S = \{\emptyset\}$, then $\eta(S, a) = \{\emptyset\}$. Otherwise,* $\eta(\{s_1, \dots, s_k\}, a) = \{s_1 + 1, \dots, s_k + 1\}$, *where the addition is modulo n.*
2. *If $\{0, 1\} \cap S = \emptyset$, then*
   (a) $\eta(S, b) = S$,
   (b) $\eta(\{0\} \cup S, b) = \{1\} \cup S$,
   (c) $\eta(\{1\} \cup S, b) = \{0\} \cup S$,
   (d) $\eta(\{0, 1\} \cup S, b) = \{0, 1\} \cup S$.
3. *If $\{0, n-1\} \cap S = \emptyset$, then*
   (a) $\eta(S, c) = \{S, \{n-1\} \cup S\}$,
   (b) $\eta(\{0, n-1\} \cup S, c) = \{\{0, n-1\} \cup S, \{0\} \cup S\}$,
   (c) $\eta(\{0\} \cup S, c) = \emptyset$,
   (d) $\eta(\{n-1\} \cup S, c) = \emptyset$.

Proof. The reverse of DFA $\mathcal{D}_n$ is the NFA $\mathcal{D}_n^{\mathbb{R}} = (Q, \Sigma, \delta^{\mathbb{R}}, \{n-1\}, \{0\})$, where $\delta^{\mathbb{R}}$ is defined by $\delta^{\mathbb{R}}(i, a) = i - 1 \bmod n$, $\delta^{\mathbb{R}}(i, b) = \delta(i, b)$, $\delta^{\mathbb{R}}(0, c) = \{0, n-1\}$, $\delta^{\mathbb{R}}(n-1, c) = \emptyset$, and $\delta^{\mathbb{R}}(i, c) = i$, for $0 < i < n-1$. After applying determinization and reversal to $\mathcal{D}_n^{\mathbb{R}}$, the claims follow by Corollary 1. □

5 Tightness of the Upper Bounds

We now show that the upper bounds derived in Section 2 are tight by proving that the atoms of the languages L_n of Definition 3 meet those bounds.

Since the states of the átomaton $\mathcal{A}_n = (\mathbf{A}, \Sigma, \eta, \mathbf{I}, \{\mathbf{A}_{m-1}\})$ are atom symbols $\mathbf{A}_i$, and the right language of each $\mathbf{A}_i$ is the atom A_i, the languages A_i are properly represented by the átomaton. Since, however, the átomaton is an NFA, to find the quotient complexity of A_i, we need the equivalent minimal DFA.

Let $\mathcal{D}_n$ be the n-state quotient DFA of Definition 3 for $n \geqslant 2$, and recall that $L(\mathcal{D}_n) = L_n$. In the sequel, using Corollary 1, we represent the átomaton $\mathcal{A}_n$ of L_n by the isomorphic NFA $\mathcal{D}_n^{\mathbb{R}\mathbb{D}\mathbb{R}} = (S, \Sigma, \alpha^{\mathbb{R}}, G, \{\mathbf{F}\})$, and identify the atoms by their sets of uncomplemented quotients. To simplify the notation, we represent atoms by the subscripts of the quotients, that is, by subsets of $Q = \{0, \dots, n-1\}$, as in Definition 3.

In this framework, to find the quotient complexity of an atom A_P, with $P \subseteq Q$, we start with the NFA $\mathcal{A}_P = (S, \Sigma, \alpha^{\mathbb{R}}, \{P\}, \{\mathbf{F}\})$, which has the same states, transitions, and final state as the átomaton, but has only one initial state, P, corresponding to the atom symbol $\mathbf{A}_P$. Because $\mathcal{A}_P^{\mathbb{R}}$ is deterministic and $\mathcal{A}_P$ has no empty states, $\mathcal{A}_P^{\mathbb{D}}$ is minimal by Theorem 2. Therefore, $\mathcal{A}_P^{\mathbb{D}}$ is the quotient

DFA of the atom A_P. The states of $\mathcal{A}_P^{\mathbb{D}}$ are certain *sets of sets* of quotient symbols; to reduce confusion we refer to them as *collections of sets*. The particular collections appearing in $\mathcal{A}_P^{\mathbb{D}}$ will be called "super-algebras".

Let U be a subset of Q with $|U| = u$, and let V be a subset of U with $|V| = v$. Define $\langle V \rangle_U$ to be the collection of all 2^{u-v} subsets of U containing V. There are $C_u^n C_v^u$ collections of the form $\langle V \rangle_U$, because there are C_u^n ways of choosing U, and for each such choice there are C_v^u ways of choosing V. The collection $\langle V \rangle_U$ is called the *super-algebra of* U *generated by* V. The *type* of a super-algebra $\langle V \rangle_U$ is the ordered pair $(|V|, |U|) = (v, u)$.

The following theorem is a well-known result of Piccard [5] about the group—known as the *symmetric group*—of all permutations of a finite set:

Theorem 5 (Permutations). *The symmetric group of size* $n!$ *of all permutations of a set* $Q = \{0, \ldots, n-1\}$ *is generated by any cyclic permutation of* Q *together with any transposition.*

Lemma 1 (Strong-Connectedness of Super-Algebras). *Super-algebras of the same type are strongly connected by words in* $\{a, b\}^*$.

Proof. Let $\langle V_1 \rangle_{U_1}$ and $\langle V_2 \rangle_{U_2}$ be any two super-algebras of the same type. Arrange the elements of V_1 in increasing order, and do the same for the elements of the sets V_2, $U_1 \setminus V_1$, $U_2 \setminus V_2$, $Q \setminus U_1$, and $Q \setminus U_2$. Let $\pi : Q \to Q$ be the mapping that assigns the ith element of V_2 to the ith element of V_1, the ith element of $U_2 \setminus V_2$ to the ith element of $U_1 \setminus V_1$, and the ith element of $Q \setminus U_2$ to the ith element of $Q \setminus U_1$. For any R_1 such that $V_1 \subseteq R_1 \subseteq U_1$, there is a corresponding subset $R_2 = \pi(R_1)$, where $V_2 \subseteq R_2 \subseteq U_2$. Thus π establishes a one-to-one correspondence between the elements of the super-algebras $\langle V_1 \rangle_{U_1}$ and $\langle V_2 \rangle_{U_2}$. Also, π is a permutation of Q, and so can be performed by a word $w \in \{a, b\}^*$ in $\mathcal{D}_n$, in view of Theorem 5. Thus every set R_2 defined as above is reachable from R_1 by w. So $\langle V_2 \rangle_{U_2}$ is reachable from $\langle V_1 \rangle_{U_1}$. □

Lemma 2 (Reachability). *Let* $\langle V \rangle_U$ *be any super-algebra of type* (v, u)*. If* $v \geqslant 2$*, then from* $\langle V \rangle_U$ *we can reach a super-algebra of type* $(v-1, u)$*. If* $u \leqslant n-2$*, then from* $\langle V \rangle_U$ *we can reach a super-algebra of type* $(v, u+1)$*.*

Proof. If $v \geqslant 2$, then by Lemma 1, from $\langle V \rangle_U$ we can reach a super-algebra $\langle V' \rangle_{U'}$ of type (v, u) such that $\{0, n-1\} \subseteq V'$. By input c we reach $\langle V' \setminus \{n-1\} \rangle_{U'}$ of type $(v-1, u)$. For the second claim, if $u \leqslant n-2$, then by Lemma 1, from $\langle V \rangle_U$ we can reach a super-algebra $\langle V' \rangle_{U'}$ of type (v, u) such that $\{0, n-1\} \cap V' = \emptyset$. By input c we reach $\langle V' \rangle_{U' \cup \{n-1\}}$ of type $(v, u+1)$. □

The next proposition holds for $n \geqslant 1$ if we let $L_1 = \Sigma^*$.

Proposition 6 (Atoms with 0 or n **Complemented Quotients)** *For* $n \geqslant 1$*, the quotient complexity of the atoms* A_Q *and* $A_\emptyset$ *of* L_n *is* $2^n - 1$.

Proof. Let $\mathcal{A}_Q$ ($\mathcal{A}_\emptyset$) be the modified átomaton with only one initial state, Q ($\emptyset$). By the considerations above, $\mathcal{A}_Q^{\mathbb{D}}$ ($\mathcal{A}_\emptyset^{\mathbb{D}}$) is the quotient DFA of A_Q ($A_\emptyset$); hence it suffices to prove the reachability of $2^n - 1$ collections.

For A_Q, the initial state of $\mathcal{A}_Q^{\mathbb{D}}$ is the collection $\{Q\}$, which is the super-algebra $\langle Q\rangle_Q$ of Q generated by Q. Now suppose that we have reached a super-algebra of type (v, n). By Lemma 1, we can reach every other super-algebra of type (v, n). If $v \geqslant 2$, then by Lemma 2 we can reach a super-algebra of type $(v-1, n)$. Thus we can reach all super-algebras $\langle V\rangle_Q$ of Q, one for each non-empty subset V of Q. Since there are at most $2^n - 1$ collections and that many can be reached, no other collection can be reached.

For $A_\emptyset$, the initial state of $\mathcal{A}_\emptyset^{\mathbb{D}}$ is the empty collection, which is the super-algebra $\langle\emptyset\rangle_\emptyset$ of $\emptyset$ generated by $\emptyset$. Now suppose we have reached a super-algebra of type $(0, u)$. By Lemma 1, we can reach every other super-algebra of type $(0, u)$. If $u \leqslant n-2$, then by Lemma 2 we can reach a super-algebra of type $(0, u+1)$. Thus we can reach all super-algebras $\langle\emptyset\rangle_U$, one for each non-empty subset U of Q. Since there are at most $2^n - 1$ collections and that many can be reached, no other collection can be reached. Hence the proposition holds. □

Proposition 7 (Tightness). *For $n \geqslant 2$ and $1 \leqslant r \leqslant n-1$, the quotient complexity of any atom of L_n with r complemented quotients is $f(n, r)$.*

Proof. Let A_P be an atom of L_n with $n-r$ uncomplemented quotients, where $1 \leqslant r \leqslant n-1$, that is, let P be the set of subscripts of the uncomplemented quotients. Let $\mathcal{A}_P$ be the modified átomaton with the initial state P. As discussed above, $\mathcal{A}_P^{\mathbb{D}}$ is minimal; hence it suffices to prove the reachability of $f(n, r)$ collections.

We start with the super-algebra $\langle P\rangle_P$ with type $(n-r, n-r)$. By Lemmas 1 and 2, we can now reach all super-algebras of types

$$\begin{gathered}(n-r, n-r), (n-r-1, n-r), \ldots, (1, n-r),\\ (n-r, n-r+1), (n-r-1, n-r+1), \ldots, (1, n-r+1),\\ \cdots\\ (n-r, n-1), (n-r-1, n-1), \ldots, (1, n-1).\end{gathered}$$

Since the number of super-algebras of type (v, u) is $C_u^n C_v^u$, we can reach

$$g(n, r) = \sum_{u=n-r}^{n-1} \sum_{v=1}^{n-r} C_u^n \cdot C_v^u$$

algebras. Changing the first summation index to $k = n - u$, we get

$$g(n, r) = \sum_{k=1}^{r} \sum_{v=1}^{n-r} C_{n-k}^n \cdot C_v^{n-k}.$$

Note that $C_{n-k}^n C_v^{n-k} = C_{k+v}^n C_k^{k+v}$, because $C_{n-k}^n C_v^{n-k} = \frac{n!}{(n-k)!k!} \cdot \frac{(n-k)!}{v!(n-k-v)!} = \frac{n!}{k!v!(n-k-v)!}$, and $C_{k+v}^n C_k^{k+v} = \frac{n!}{(k+v)!(n-k-v)!} \cdot \frac{(k+v)!}{k!v!} = \frac{n!}{(n-k-v)!k!v!}$. Now, we can write $g(n, r) = \sum_{k=1}^{r} \sum_{v=1}^{n-r} C_{k+v}^n \cdot C_k^{k+v}$, and changing the second summation index to $h = k + v$, we have

$$g(n, r) = \sum_{k=1}^{r} \sum_{h=k+1}^{k+n-r} C_h^n \cdot C_k^h.$$

We notice that $g(n, r) = f(n, r) - 1$. From the super-algebra $\langle V \rangle_V$, where $V = \{0, 1, \ldots, n - r - 1\}$, we reach the empty quotient by input c, since V contains 0, but not $n - 1$.

Since we can reach $f(n, r)$ super-algebras, no other collection can be reached, and the proposition holds. □

6 Conclusions

The atoms of a regular language L are its basic building blocks. We have studied the quotient complexity of the atoms of L as a function of the quotient complexity of L. We have computed an upper bound for the quotient complexity of any atom with r complemented quotients, and exhibited a class $\{L_n\}$ of languages whose atoms meet this bound.

Acknowledgments. We are grateful to Baiyu Li for writing a program for evaluating the quotient complexity of atoms. We thank Eric Rowland and Jeff Shallit for computing the ratio defined for Table 1 for some large values of n.

References

1. Brzozowski, J.: Canonical regular expressions and minimal state graphs for definite events. In: Proceedings of the Symposium on Mathematical Theory of Automata. MRI Symposia Series, vol. 12, pp. 529–561. Polytechnic Press, Polytechnic Institute of Brooklyn, N.Y. (1963)
2. Brzozowski, J.: Quotient complexity of regular languages. J. Autom. Lang. Comb. 15(1/2), 71–89 (2010)
3. Brzozowski, J., Tamm, H.: Theory of Átomata. In: Mauri, G., Leporati, A. (eds.) DLT 2011. LNCS, vol. 6795, pp. 105–116. Springer, Heidelberg (2011)
4. Dénes, J.: On transformations, transformation semigroups and graphs. In: Erdös, P., Katona, G. (eds.) Theory of Graphs. Proceedings of the Colloquium on Graph Theory held at Tihany, 1966, pp. 65–75. Akadémiai Kiado (1968)
5. Piccard, S.: Sur les fonctions définies dans les ensembles finis quelconques. Fund. Math. 24, 298–301 (1935)
6. Salomaa, A., Wood, D., Yu, S.: On the state complexity of reversals of regular languages. Theoret. Comput. Sci. 320, 315–329 (2004)
7. Yu, S.: State complexity of regular languages. J. Autom. Lang. Comb. 6, 221–234 (2001)