Most Complex Regular Right-Ideal Languages*

Janusz Brzozowski[1] and Gareth Davies[2]

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

Abstract. A right ideal is a language L over an alphabet Σ that satisfies the equation $L = L\Sigma^*$. We show that there exists a sequence $(R_n \mid n \geqslant 3)$ of regular right-ideal languages, where R_n has n left quotients and is most complex among regular right ideals under the following measures of complexity: the state complexities of the left quotients, the number of atoms (intersections of complemented and uncomplemented left quotients), the state complexities of the atoms, the size of the syntactic semigroup, the state complexities of reversal, star, product, and all binary boolean operations that depend on both arguments. Thus $(R_n \mid n \geqslant 3)$ is a universal witness reaching the upper bounds for these measures.

Keywords: atom, operation, quotient, regular language, right ideal, state complexity, syntactic semigroup, universal witness.

1 Introduction

Brzozowski [3] called a regular language *most complex* if it meets the upper bounds for a large set of commonly used language properties and operations, and found a single *witness* language of state complexity n for each $n \geqslant 3$ that meets all these bounds. In particular, this language has the maximal number of atoms and the state complexities of these atoms are maximal. Moreover, it meets the upper bounds for the state complexities of all the basic operations: reverse, Kleene star, boolean operations, product (also known as concatenation or catenation), as well as a large number of combined operations. In view of this, such a witness has been called *universal*.

If we restrict our attention to some subclass of regular languages, then the universal witness mentioned above no longer works because it lacks the properties of the subclass. In this paper we ask whether the approach used for general regular languages can be extended to subclasses. We answer this question positively for regular right ideals by presenting a universal right-ideal witness.

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

H. Jürgensen et al. (Eds.): DCFS 2014, LNCS 8614, pp. 90–101, 2014.

For a further discussion of regular right ideals see [5,8]. It was pointed out in [5] that right ideals deserve to be studied for several reasons: They are fundamental objects in semigroup theory, they appear in the theoretical computer science literature as early as 1965, and they continue to be of interest in the present. Right ideal languages are complements of prefix-closed languages. Besides being of theoretical interest, right ideals also play a role in algorithms for pattern matching: When searching for all words beginning in a word from some set L, one is looking for all the words of the right ideal $L\Sigma^*$.

2 Background

A *deterministic finite automaton (DFA)* $\mathcal{D} = (Q, \Sigma, \delta, q_1, F)$ consists of a finite non-empty set Q of *states*, a finite non-empty *alphabet* Σ, a *transition function* $\delta\colon Q \times \Sigma \to Q$, an *initial state* $q_1 \in Q$, and a set $F \subseteq Q$ of *final* states. The transition function is extended to functions $\delta'\colon Q \times \Sigma^* \to Q$ and $\delta''\colon 2^Q \times \Sigma^* \to 2^Q$ as usual, and these extensions are also denoted by δ. A state q of a DFA is *reachable* if there is a word $w \in \Sigma^*$ such that $\delta(q_1, w) = q$. The *language accepted* by $\mathcal{D}$ is $L(\mathcal{D}) = \{w \in \Sigma^* \mid \delta(q_1, w) \in F\}$. The *language of a state* q is the language accepted by the DFA $\mathcal{D}_q = (Q, \Sigma, \delta, q, F)$. A state is *empty* if its language is empty. Two DFAs are *equivalent* if their languages are the same. Two states are *equivalent* if their languages are equal; otherwise, they are *distinguishable* by some word that is in the language of one of the states, but not of the other. If $S \subseteq Q$, two states $p, q \in Q$ are *distinguishable with respect to* S if there is a word w such that $\delta(p, w) \in S$ if and only if $\delta(q, w) \notin S$. A DFA is *minimal* if all of its states are reachable and no two states are equivalent.

A *nondeterministic finite automaton (NFA)* is a tuple $\mathcal{N} = (Q, \Sigma, \eta, Q_1, F)$, where Q, Σ, and F are as in a DFA, $\eta\colon Q \times \Sigma \to 2^Q$ is the transition function and $Q_1 \subseteq Q$ is the *set of initial states*. An *ε-NFA* has all the features of an NFA but its transition function $\eta\colon Q \times (\Sigma \cup \{\varepsilon\}) \to 2^Q$ allows also transitions under the empty word ε. The *language accepted* by an NFA or an ε-NFA is the set of words w for which there exists a sequence of transitions such that the concatenation of the symbols inducing the transitions is w, and this sequence leads from a state in Q_1 to a state in F. Two NFAs are *equivalent* if they accept the same language.

We use the following operations on automata:

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

Let $\mathcal{D} = (Q, \Sigma, \delta, q_1, F)$ be a DFA. For each word $w \in \Sigma^*$, the transition function induces a transformation t_w of Q by w: for all $q \in Q$, $qt_w \stackrel{\text{def}}{=} \delta(q, w)$. The set $T_\mathcal{D}$ of all such transformations by non-empty words forms a semigroup of transformations called the *transition semigroup* of $\mathcal{D}$ [11]. Conversely, we can use a set $\{t_a \mid a \in \Sigma\}$ of transformations to define δ, and so also the DFA $\mathcal{D}$. We also write $a\colon t$ to mean that the transformation induced by $a \in \Sigma$ is t.

The *syntactic congruence* $\leftrightarrow_L$ of a language $L \subseteq \Sigma^*$ is defined on Σ^+: For $x, y \in \Sigma^+$, $x \leftrightarrow_L y$ if and only if $uxv \in L \Leftrightarrow uyv \in L$ for all $u, v \in \Sigma^*$. The quotient set $\Sigma^+/\leftrightarrow_L$ of equivalence classes of the relation $\leftrightarrow_L$ is a semigroup called the *syntactic semigroup* of L. If $\mathcal{D}$ is the minimal DFA of L, then $T_{\mathcal{D}}$ is isomorphic to the syntactic semigroup T_L of L [11], and we represent elements of T_L by transformations in $T_{\mathcal{D}}$.

A *permutation* of Q is a mapping of Q *onto* itself. The *identity* transformation $\mathbf{1}$ maps each element to itself, that is, $q\mathbf{1} = q$ for $q \in Q$. A transformation t is a *cycle* of length k if there exist pairwise different elements $p_1, \ldots, p_k$ such that $p_1 t = p_2, p_2 t = p_3, \ldots, p_{k-1} t = p_k, p_k t = p_1$, and other elements of Q are mapped to themselves. A cycle is denoted by $(p_1, p_2, \ldots, p_k)$. A *transposition* is a cycle (p, q). For $p \neq q$, a *unitary* transformation $t\colon (p \to q)$, has $pt = q$ and $rt = r$ for all $r \neq p$.

The set of all permutations of a set Q of n elements is a group, called the *symmetric group* of degree n. Without loss of generality, from now on we assume that $Q = \{1, 2, \ldots, n\}$. It is well known that the symmetric group of degree n can be generated by any cyclic permutation of n elements together with any transposition. In particular, it can be generated by $(1, 2, \ldots, n)$ and $(1, 2)$.

The set of all transformations of a set Q, denoted by $\mathcal{T}_Q$, is a monoid with $\mathbf{1}$ as the identity. It is well known that the transformation monoid $\mathcal{T}_Q$ of size n^n can be generated by any cyclic permutation of n elements together with any transposition and any unitary transformation. In particular, $\mathcal{T}_Q$ can be generated by $(1, 2, \ldots, n)$, $(1, 2)$ and $(n \to 1)$.

The *state complexity* [12] *of a regular language* L over an alphabet Σ is the number of states in any minimal DFA recognizing L. An equivalent notion is that of *quotient complexity* [2], which is the number of distinct left quotients of L, where the left quotient of $L \subseteq \Sigma^*$ by a word $w \in \Sigma^*$ is the language $w^{-1}L = \{x \in \Sigma^* \mid wx \in L\}$. This paper uses *complexity* for both of these equivalent notions, and this term will not be used for any other property here.

The *(state/quotient) complexity of an operation* [12] on regular languages is the maximal complexity of the language resulting from the operation as a function of the complexities of the arguments. For example, for $L \subseteq \Sigma^*$, the complexity of the reverse L^R of L is 2^n if the complexity of L is n, since a minimal DFA for L^R can have at most 2^n states and there exist languages meeting this bound [9].

There are two parts to the process of establishing the complexity of an operation. First, one must find an *upper bound* on the complexity of the result of the operation by using quotient computations or automaton constructions. Second, one must find *witnesses* that meet this upper bound. One usually defines a sequence $(L_n \mid n \geqslant k)$ of languages, where k is some small positive integer. This sequence will be called a *stream*. The languages in a stream differ only in the parameter n. For example, one might study unary languages $(\{a^n\}^* \mid n \geqslant 1)$ that have zero occurrences of the letter a modulo n. A unary operation takes its argument from a stream $(L_n \mid n \geqslant k)$. For a binary operation, one adds a stream $(K_n \mid n \geqslant k)$ as the second argument. While the witness streams are normally

different for different operations, our main result shows that a single stream can meet the complexity bounds for all operations in the case of right ideals.

Atoms of regular languages were studied in [7], and their complexities, in [6]. Let L be a regular language with quotients $K = \{K_1, \ldots, K_n\}$. Each subset S of K defines an *atomic intersection* $A_S = \widetilde{K_1} \cap \cdots \cap \widetilde{K_n}$, where $\widetilde{K_i}$ is K_i if $K_i \in S$ and $\overline{K_i}$ otherwise. An *atom* of L is a non-empty atomic intersection. Since non-empty atomic intersections 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 and is denoted by $\mathcal{B}(A)$. The *cobasis* of A is $\overline{\mathcal{B}}(A) = K \setminus \mathcal{B}(A)$. The basis of an atom is the set of quotients of L that occur uncomplemented as terms of the corresponding intersection, and the cobasis is the set of quotients that occur complemented.

It was proven in [7] that each regular language L defines a unique set of atoms, that every quotient of L (including L itself) is a union of atoms, and that every quotient of every atom of L is a union of atoms. Thus the atoms of L are its basic building blocks. It was argued in [3] that the complexity of the atoms of a language should be considered when searching for "most complex" regular languages, since a complex language should have complex building blocks. We shall show that – as was the case for arbitrary regular languages – for right ideals there is a tight upper bound on the complexity of any atom with a basis of a given size.

3 Main Results

The stream of right ideals that turns out to be most complex is defined as follows:

Definition 1. *For $n \geqslant 3$, let $\mathcal{R}_n = \mathcal{R}_n(a,b,c,d) = (Q, \Sigma, \delta, 1, \{n\})$, where $Q = \{1, \ldots, n\}$ is the set of states*[1]*, $\Sigma = \{a,b,c,d\}$ is the alphabet, the transformations defined by δ are $a\colon (1, \ldots, n-1)$, $b\colon (2, \ldots, n-1)$, $c\colon (n-1 \to 1)$ and $d\colon (n-1 \to n)$, 1 is the initial state, and $\{n\}$ is the set of final states. Let $R_n = R_n(a,b,c,d)$ be the language accepted by $\mathcal{R}_n$.*

The structure of the DFA $\mathcal{R}_n(a,b,c,d)$ is shown in Figure 1. Note that input b induces the identity transformation in $\mathcal{R}_n$ for $n = 3$.

The stream of languages of Definition 1 is very similar to the stream $(L_n \mid n \geqslant 2)$ shown to be a universal witness for regular languages in [3,6]. In that stream, L_n is defined by the DFA $\mathcal{D}_n = \mathcal{D}_n(a,b,c) = (Q, \Sigma, \delta, 1, \{n\})$, where $Q = \{1, \ldots, n\}$, $\Sigma = \{a,b,c\}$, and δ is defined by $a\colon (1, \ldots, n)$, $b\colon (1,2)$, and $c\colon (n \to 1)$. The automaton $\mathcal{R}_n$ can be constructed by taking $\mathcal{D}_{n-1}$, adding a new state n and a new input $d\colon (n-1 \to n)$, making n the only final state, and having b induce the cyclic permutation $(2, \ldots, n-1)$, rather than the transposition $(1,2)$. The new state and input are necessary to ensure that R_n is a right ideal for all n.

[1] Although Q and δ depend on n, this dependence is usually not shown to keep the notation as simple as possible.

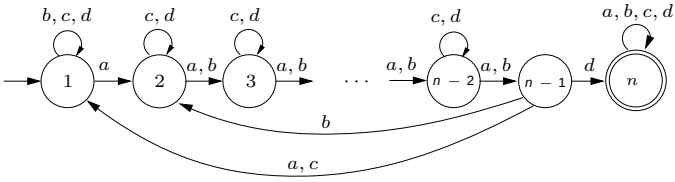

Fig. 1. Automaton $\mathcal{R}_n$ of a most complex right ideal R_n

Changing the transformation induced by b is necessary since, if b induces $(1, 2)$ in $\mathcal{R}_n$, then R_n does not meet the bound for product.

We can generalize this definition to a stream $(R_n \mid n \geqslant 1)$ by noting that when $n = 1$, all four inputs induce the identity transformation, and when $n = 2$, a, b and c induce the identity transformation, while d induces $(1 \to 2)$. Hence $R_1 = \{a, b, c, d\}^*$ and $R_2 = \{a, b, c\}^* d \{a, b, c, d\}^*$. However, the complexity bound for star is not reached by R_1, and the complexity bounds for boolean operations are not reached when one of the operands is R_1 or R_2. Thus we require $n \geqslant 3$.

In some cases, the complexity bounds can be reached even when the alphabet size is reduced. If c is not needed, let $\mathcal{R}_n(a, b, d)$ be the DFA of Definition 1 restricted to inputs a, b and d, and let $R_n(a, b, d)$ be the language recognized by this DFA. If both b and c are not needed, we use $\mathcal{R}_n(a, d)$ and $R_n(a, d)$. We also define $\mathcal{R}_n(b, a, d)$ to be the DFA obtained from $\mathcal{R}_n(a, b, d)$ by interchanging the roles of the inputs a and b, and let $R_n(b, a, d)$ be the corresponding language.

Theorem 1 (Main Results). *The language $R_n = R_n(a, b, c, d)$ has the properties listed below. Moreover, all the complexities of R_n are the maximal possible for right ideals. The results hold for all $n \geqslant 1$ unless otherwise specified.*

1. *$R_n(a, d)$ has n quotients, that is, its (state/quotient) complexity is n.*
2. *The syntactic semigroup of $R_n(a, b, c, d)$ has cardinality n^{n-1}.*
3. *Quotients of $R_n(a, d)$ have complexity n, except for the quotient $\{a, d\}^*$, which has complexity 1.*
4. *$R_n(a, b, c, d)$ has 2^{n-1} atoms.*
5. *The atom of $R_n(a, b, c, d)$ with the empty cobasis has complexity 2^{n-1}. If an atom of $R_n(a, b, c, d)$ has a cobasis of size r, $1 \leqslant r \leqslant n - 1$, its complexity is*

$$1 + \sum_{k=1}^{r} \sum_{h=k+1}^{k+n-r} \binom{n-1}{h-1} \binom{h-1}{k}.$$

6. *The reverse of $R_n(a, d)$ has complexity 2^{n-1}.*
7. *For $n \geqslant 2$, the star of $R_n(a, d)$ has complexity $n + 1$.*
8. *For $m, n \geqslant 3$, the complexity of $R_m(a, b, d) \cap R_n(b, a, d)$ is mn.*
9. *For $m, n \geqslant 3$, the complexity of $R_m(a, b, d) \oplus R_n(b, a, d)$ is mn.*
10. *For $m, n \geqslant 3$, the complexity of $R_m(a, b, d) \setminus R_n(b, a, d)$ is $mn - (m - 1)$.*
11. *For $m, n \geqslant 3$, the complexity of $R_m(a, b, d) \cup R_n(b, a, d)$ is $mn - (m + n - 2)$.*

12. *For $m, n \geqslant 3$, since any binary boolean operation can be expressed as a combination of the four operations above (and complement, which does not affect complexity), the complexity of $R_m(a, b, d) \circ R_n(b, a, d)$ is maximal for all binary boolean operations $\circ$.*
13. *For $m, n \geqslant 3$, if $m \neq n$, then the complexity of $R_m(a, b, d) \circ R_n(a, b, d)$ is maximal for all binary boolean operations $\circ$.*
14. *The complexity of $R_m(a, b, d) \cdot R_n(a, b, d)$ is $m + 2^{n-2}$.*

The proof of Theorem 1 is the topic of the remainder of the paper.

4 Conditions for the Complexity of Right Ideals

1. Complexity of the Language: $R_n(a, d)$ has n quotients because the DFA $\mathcal{R}_n(a, d)$ is minimal. This holds since the non-final state i accepts $a^{n-1-i}d$ and no other non-final state accepts this word, for $1 \leqslant i \leqslant n-1$, and all non-final states are distinguishable from the final state n by the empty word.

2. Cardinality of the Syntactic Semigroup: It was proved in [8] that the syntactic semigroup of a right ideal of complexity n has cardinality at most n^{n-1}. To show $R_n(a, b, c, d)$ meets this bound, one first verifies the following:

Remark 1. For $n \geqslant 3$, the transposition $(1, 2)$ in $\mathcal{R}_n$ is induced by $a^{n-2}b$.

Theorem 1 (2) The syntactic semigroup of $R_n(a, b, c, d)$ has cardinality n^{n-1}.

Proof. The cases $n \leqslant 3$ are easily checked. For $n \geqslant 4$, let the DFA $\mathcal{P}_n$ be $\mathcal{P}_n = (Q, \Sigma, \delta, 1, \{n\})$, where $Q = \{1, \ldots, n\}$, $\Sigma = \{a, b, c, d\}$, and $a\colon (1, \ldots, n-1)$, $b\colon (1, 2)$, $c\colon (n-1 \to 1)$ and $d\colon (n-1 \to n)$. It was proved in [8] that the syntactic semigroup of $P_n(a, b, c, d)$ has cardinality n^{n-1}. Since words in Σ^* can induce all the transformations of $\mathcal{P}_n$ in $\mathcal{R}_n(a, b, c, d)$, the claim follows. □

3. Complexity of Quotients: Each quotient of $R_n(a, d)$, except the quotient $\{a, d\}^*$, has complexity n, since states $1, \ldots, n-1$ are strongly connected. So the complexities of the quotients are maximal for right ideals.

4. Number of Atoms: It was proved in [6] that the number of atoms of L is precisely the complexity of the reverse of L. It was shown in [5] that the maximal complexity of L^R for right ideals is 2^{n-1}. For $n \leqslant 3$ it is easily checked that our witness meets this bound. For $n > 3$, it was proved in [8] that the reverse of $R_n(a, d)$, and hence also of $R_n(a, b, c, d)$, reaches this bound.

5. Complexity of Atoms: This is the topic of Section 5.

6. Reversal: See **4. Number of Atoms**.

7. Star: The complexity of the star of a right ideal is at most $n + 1$ [5]. This follows because, if $\varepsilon \notin L$, we need to add ε to $L = L\Sigma^*$ to obtain L^*. Our witness meets this bound, as one can easily verify:

Remark 2 (Star). For $n \geqslant 2$, the complexity of $(R_n(a, d))^*$ is $n + 1$.

8.–14. Boolean Operations and **Product:** See Sections 6 and 7.

Table 1. Maximal complexity of atoms of right ideals

n	1	2	3	4	5	6	7	$\cdots$
r=0	**1/1**	**2/3**	4/7	8/15	16/31	32/63	64/127	$\cdots$
r=1		**2/3**	**5/10**	13/29	33/76	81/187	193/442	$\cdots$
r=2		$*$/**3**	4/**10**	**16/43**	**53/141**	156/406	427/1,086	$\cdots$
r=3			$*$/7	8/29	43/**141**	**166/501**	**542/1,548**	$\cdots$
r=4				$*$/15	16/76	106/406	462/**1,548**	$\cdots$
r=5					$*$/31	32/187	249/1,086	$\cdots$
r=6						$*$/63	64/442	$\cdots$
max	1/1	2/3	5/10	16/43	53/141	166/501	542/1,548	$\cdots$
ratio	–	2/3	2.50/3.33	3.20/4.30	3.31/3.28	3.13/3.55	3.27/3.09	$\cdots$

5 Complexity of Atoms

In [6], for the language stream $(L_n \mid n \geqslant 2)$ described after Definition 1, it was proved that the atoms of L_n have maximal complexity amongst all regular languages of complexity n. We want to prove that the atoms of $R_n(a,b,c,d)$ have maximal complexity amongst all right ideals of complexity n. We only give a high-level outline following the approach of [6].

1. Derive upper bounds for the complexities of atoms of right ideals.
The cobasis of an atom cannot contain Σ^*; if it did, then $\overline{\Sigma^*} = \emptyset$ would be a term in the corresponding atomic intersection and the intersection would be empty. Since all right ideals have Σ^* as a quotient, every atom of a right ideal must contain Σ^* in its basis. It follows the cobasis of an atom of a right ideal is either empty or contains r quotients, where $1 \leqslant r \leqslant n-1$. Knowing this, we can derive the upper bounds by the same method as in [6].

2. Describe the transition function of the átomaton of $R_n(a,b,c,d)$.
Let $\mathbf{A} = \{A_1, \ldots, A_m\}$ be the set of atoms of L. The *átomaton*[2] of L is the NFA $\mathcal{A} = (\mathbf{A}, \Sigma, \eta, \mathbf{A}_I, A_f)$, where the *initial* atoms are $\mathbf{A}_I = \{A_i \mid L \in \mathcal{B}(A_i)\}$, the *final* atom A_f is the unique atom such that $K_i \in \mathcal{B}(A_f)$ if and only if $\varepsilon \in K_i$, and $A_j \in \eta(A_i, a)$ if and only if $aA_j \subseteq A_i$. In the átomaton the language of state A of $\mathcal{A}$ is the atom A of L. Since each regular language defines a unique set of atoms, each regular language also defines a unique átomaton.

3. Prove that certain strong connectedness and reachability results hold for states of minimal DFAs of atoms of $R_n(a,b,c,d)$.

4. Prove that the complexity of each atom of $R_n(a,b,c,d)$ meets the established bound.

Many steps of this proof are similar or identical to the proof for L_n given in [6]; for the details see [4]. Table 4 shows the bounds for right ideals (first entry) and compares them to those of regular languages (second entry). An asterisk indicates the case is impossible for right ideals. The *ratio* row shows the ratio m_n/m_{n-1} for $n \geqslant 2$, where m_i is the i^{th} entry in the *max* row.

[2] The accent in *átomaton* avoids confusion with *automaton*, and suggests that the stress should be on the first syllable, since the word comes from *atom*.

6 Boolean Operations

Tight upper bounds for boolean operations on right ideals [5] are mn for intersection and symmetric difference, $mn-(m+n)$ for difference, and $mn-(m+n-2)$ for union. Since $L_n \cup L_n = L_n \cap L_n = L_n$, and $L_n \setminus L_n = L_n \oplus L_n = \emptyset$, two different languages must be used to reach the bounds if $m = n$. We use $R_m = R_m(a, b, d)$ and $R_n = R_n(b, a, d)$, shown in Figure 2 for $m = 4$ and $n = 5$.

Let $\mathcal{R}_{m,n} = \mathcal{R}_m \times \mathcal{R}_n = (Q_m \times Q_n, \Sigma, \delta, (1,1), F_{m,n})$ with $\delta((i,j),\sigma) = (\delta_m(i,\sigma), \delta_n(j,\sigma))$, where δ_m (δ_n) is the transition function of $\mathcal{R}_m$ ($\mathcal{R}_n$). Depending on $F_{m,n}$, this DFA recognizes different boolean operations on R_m and R_n. The direct product of $\mathcal{R}_4(a,b,d)$ and $\mathcal{R}_5(b,a,d)$ is in Figure 3.

In our proof that the bounds for boolean operations are reached, we use a result of Bell, Brzozowski, Moreira and Reis [1]. A binary boolean operation $\circ$ on regular languages is a mapping $\circ : 2^{\Sigma^*} \times 2^{\Sigma^*} \to 2^{\Sigma^*}$. If $L, L' \subseteq \Sigma^*$, the result of the operation $\circ$ is denoted by $L \circ L'$. We say that such a boolean operation is *proper* if $\circ$ is not a constant, and not a function of one variable only, that is, it is not the identity or the complement of one of the variables.

Let S_n denote the symmetric group of degree n. A *basis* [10] of S_n is an ordered pair (s,t) of distinct transformations of $Q_n = \{1,\ldots,n\}$ that generate S_n. Two bases (s,t) and (s',t') of S_n are *conjugate* if there exists a transformation $r \in S_n$ such that $rsr^{-1} = s'$, and $rtr^{-1} = t'$. A DFA *has a basis* (t_a, t_b) *for* S_n if it has letters $a, b \in \Sigma$ such that a induces t_a and b induces t_b.

Proposition 1 (Symmetric Groups and Boolean Operations [1]). *Suppose that $m, n \geqslant 1$, L_m and L'_n are regular languages of complexity m and n respectively, and $\mathcal{D}_m = (Q_m, \Sigma, \delta, 1, F)$ and $\mathcal{D}'_n = (Q_n, \Sigma, \delta', 1, F')$ are minimal DFAs for L_m and L'_n, where $\emptyset \subsetneq F \subsetneq Q_m$ and $\emptyset \subsetneq F' \subsetneq Q_n$. Suppose further that D_m has a basis $B = (t_a, t_b)$ for S_m and D'_n has a basis $B' = (t'_a, t'_b)$ for S_n. Let $\circ$ be a proper binary boolean operation. Then the following hold:*

1. In the direct product $\mathcal{D}_m \times \mathcal{D}'_n$, all mn states are reachable if and only if $m \neq n$, or $m = n$ and the bases B and B' are not conjugate.

2. For $m, n \geqslant 2$, $(m,n) \notin \{(2,2),(3,4),(4,3),(4,4)\}$, $L_m \circ L'_n$ has complexity mn if and only if $m \neq n$, or $m = n$ and the bases B and B' are not conjugate.

This implies that if the conditions of the proposition hold, then no matter how we choose the sets F and F', as long as $\emptyset \subsetneq F \subsetneq Q_m$ and $\emptyset \subsetneq F' \subsetneq Q_n$,

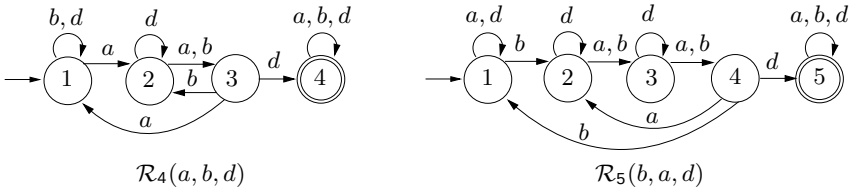

Fig. 2. Right-ideal witnesses for boolean operations

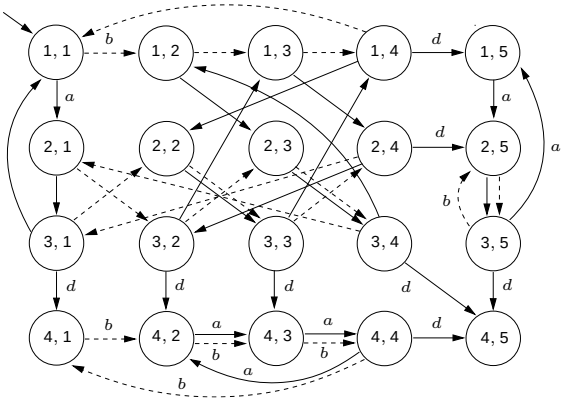

Fig. 3. Direct-product automaton for boolean operations, $m = 4, n = 5$. Transitions under a and d are in solid lines and under b, in dotted lines. Unlabelled solid transitions are under a. Self-loops are omitted.

and the boolean function $\circ$ is proper, the direct product DFA $\mathcal{D}_m \times \mathcal{D}_n$ has mn states and is minimal.

In the case of our right ideal $\mathcal{R}_m$ ($\mathcal{R}_n$), the transitions t_a and t_b (t'_a and t'_b) restricted to $\{1, \ldots, n-1\}$, constitute a basis for S_{m-1} (S_{n-1}). This implies that in the direct product $\mathcal{R}_{m,n}$, all states in the set $S = \{(i,j) \mid 1 \leqslant i \leqslant m-1, 1 \leqslant j \leqslant n-1\}$ are reachable by words in $\{a,b\}^*$. Furthermore, if $m, n \geqslant 3$ and $(m,n) \notin \{(3,3),(4,5),(5,4),(5,5)\}$, then every pair of states in S is distinguishable with respect to $F \circ F'$, the set of final states of the direct product.

Theorem 1 (8–11) (Boolean Operations) If $m, n \geqslant 3$, then

- The complexity of $R_m(a,b,d) \cap R_n(b,a,d)$ is mn.
- The complexity of $R_m(a,b,d) \oplus R_n(b,a,d)$ is mn.
- The complexity of $R_m(a,b,d) \setminus R_n(b,a,d)$ is $mn - (m-1)$.
- The complexity of $R_m(a,b,d) \cup R_n(b,a,d)$ is $mn - (m+n-2)$.

Proof. In the cases where $(m,n) \in \{(3,3),(4,5),(5,4),(5,5)\}$, we cannot apply Proposition 1, but we have verified computationally that the bounds are met. For the remainder of the proof we assume $(m,n) \notin \{(3,3),(4,5),(5,4),(5,5)\}$.

Our first task is to show that all mn states of $\mathcal{R}_{m,n}$ are reachable. By Proposition 1, all states in the set $S = \{(i,j) \mid 1 \leqslant i \leqslant m-1, 1 \leqslant j \leqslant n-1\}$ are reachable. The remaining states are the ones in the last row or last column (that is, row m or column n) of the direct product.

For $1 \leqslant j \leqslant n-2$, from state $(m-1,j)$ we can reach (m,j) by d. From state $(m,n-2)$ we can reach $(m,n-1)$ by a. From state $(m-1,n-1)$ we can reach (m,n) by d. Hence all states in row m are reachable.

For $1 \leqslant i \leqslant m-2$, from state $(i,n-1)$ we can reach (i,n) by d. From state $(m-2,n)$ we can reach $(m-1,n)$ by a. Hence all states in column n are reachable, and thus all mn states are reachable.

We now count the number of distinguishable states for each operation. Let $H = \{(m, j) \mid 1 \leqslant j \leqslant n\}$ be the set of states in the last row and let $V = \{(i, n) \mid 1 \leqslant i \leqslant m\}$ be the set of states in the last column. If $\circ \in \{\cap, \oplus, \setminus, \cup\}$, then $R_m(a, b, d) \circ R_n(b, a, d)$ is recognized by $\mathcal{R}_{m,n}$, where the set of final states is taken to be $H \circ V$.

Let $H' = \{(m-1, j) \mid 1 \leqslant j \leqslant n-1\}$ and let $V' = \{(i, n-1) \mid 1 \leqslant i \leqslant m-1\}$. By Proposition 1, all states in S are distinguishable with respect to $H' \cap V' = \{(m-1, n-1)\}$. We claim that they are also distinguishable with respect to $H \circ V$ for $\circ \in \{\cap, \oplus, \setminus, \cup\}$.

Distinguishability with respect to $H' \cap V'$ implies that for all pairs of states $(i, j), (k, \ell) \in S$, there exists a word w that sends (i, j) to $(m-1, n-1)$ and sends (k, ℓ) to some other state in S. It follows that the word wd sends (i, j) to (m, n) (which is in $H \cap V$), while (k, ℓ) is sent to a state outside of $H \cap V$. Hence all states in S are distinguishable with respect to $H \cap V$. The same argument works for $H \oplus V$, $H \setminus V$, and $H \cup V$.

Thus for each boolean operation $\circ$, all $(m-1)(n-1) = mn - m - n + 1$ states in S are distinguishable with respect to the final state set $H \circ V$. To show that the complexity bounds are reached by $R_m(a, b, d) \circ R_n(b, a, d)$, it suffices to consider how many of the $m + n - 1$ states in $H \cup V$ are distinguishable with respect to $H \circ V$.

Intersection: Here the set of final states is $H \cap V = \{(m, n)\}$. State (m, n) is the only final state and hence is distinguishable from all the other states. Any two states in H (V) are distinguished by words in b^*d (a^*d). State $(m, 1)$ accepts $b^{n-2}d$, while $(1, n)$ rejects it. For $2 \leqslant i \leqslant n-1$, (m, i) is sent to $(m, 1)$ by b^{n-1-i}, while state $(1, n)$ is not changed by that word. Hence (m, i) is distinguishable from $(1, n)$. By a symmetric argument, (j, n) is distinguishable from $(m, 1)$ for $2 \leqslant j \leqslant m-1$. For $2 \leqslant i \leqslant n-1$ and $2 \leqslant j \leqslant m-1$, (m, i) is distinguished from (j, n) because b^{n-i} sends the former to $(m, 1)$ and the latter to a state of the form (k, n), where $2 \leqslant k \leqslant m-1$. Hence all pairs of states from $H \cup V$ are distinguishable. Siince there are $m + n - 1$ states in $H \cup V$, it follows there are $(mn - m - n + 1) + (m + n - 1) = mn$ distinguishable states.

Symmetric Difference: Here the set of final states is $H \oplus V$, that is, all states in the last row and column except (m, n), which is the only empty state. This situation is complementary to that for intersection. Thus every two states from $H \cup V$ are distinguishable by the same word as for intersection. Hence there are mn distinguishable states.

Difference: Here the set of final states is $H \setminus V$, that is, all states in the last row H except (m, n), which is empty. All other states in the last column V are also empty. The m empty states in V are all equivalent, and the $n - 1$ final states in $H \setminus V$ are distinguished in the same way as for intersection. Hence there are $(n - 1) + 1 = n$ distinguishable states in $H \setminus V$. It follows there are $(mn - m - n + 1) + n = mn - (m - 1)$ distinguishable states.

Union: Here the set of final states is $H \cup V$. From a state in $H \cup V$ it is only possible to reach other states in $H \cup V$, and all these states are final; so every

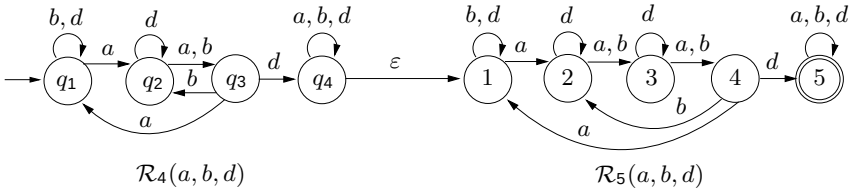

Fig. 4. Right-ideal witnesses for product

state in $H \cup V$ accepts Σ^*. Thus all the states in $H \cup V$ are equivalent, and so there are $(mn - m - n + 1) + 1 = mn - (m + n - 2)$ distinguishable states. □

Although it is impossible for the stream $(R_n(a,b,d) \mid n \geqslant 3)$ to meet the bound for boolean operations when $m = n$, this stream is as complex as it could possibly be in view of the following theorem proved in [4]:

Theorem 1 (13) (Boolean Operations, $m \neq n$) If $m, n \geqslant 3$ and $m \neq n$,

- The complexity of $R_m(a,b,d) \cap R_n(a,b,d)$ is mn.
- The complexity of $R_m(a,b,d) \oplus R_n(a,b,d)$ is mn.
- The complexity of $R_m(a,b,d) \setminus R_n(a,b,d)$ is $mn - (m-1)$.
- The complexity of $R_m(a,b,d) \cup R_n(a,b,d)$ is $mn - (m+n-2)$.

7 Product

We show that the complexity of the product of $R_m(a,b,d)$ with $R_n(a,b,d)$ reaches the maximum possible bound derived in [5]. To avoid confusing states of the two DFAs, we label their states differently. Let $\mathcal{R}_m = \mathcal{R}_m(a,b,d) = (Q'_m, \Sigma, \delta', q_1, \{q_m\})$, where $Q'_m = \{q_1, \ldots, q_m\}$, and let $\mathcal{R}_n = \mathcal{R}_n(a,b,d)$, as in Definition 1. Define the ε-NFA $\mathcal{P} = (Q'_m \cup Q_n, \Sigma, \delta_{\mathcal{P}}, \{q_1\}, \{n\})$, where $\delta_{\mathcal{P}}(q,a) = \{\delta'(q,a)\}$ if $q \in Q'_m$, $a \in \Sigma$, $\delta_{\mathcal{P}}(q,a) = \{\delta(q,a)\}$ if $q \in Q_n$, $a \in \Sigma$, and $\delta_{\mathcal{P}}(q_m, \varepsilon) = \{1\}$. This ε-NFA accepts $R_m R_n$, and is illustrated in Figure 4.

Theorem 1 (14) (Product) For $m \geqslant 1$, $n \geqslant 2$, the complexity of the product $R_m(a,b,d) \cdot R_n(a,b,d)$ is $m + 2^{n-2}$.

Proof. It was shown in [5] that $m + 2^{n-2}$ is an upper bound on the complexity of the product of two right ideals. To prove this bound is met, we apply the subset construction to $\mathcal{P}$ to obtain a DFA $\mathcal{D}$ for $R_m R_n$. The states of $\mathcal{D}$ are subsets of $Q'_m \cup Q_n$. We prove that all states of the form $\{q_i\}$, $i = 1, \ldots, m-1$ and all states of the form $\{q_m, 1\} \cup S$, where $S \subseteq Q_n \setminus \{1, n\}$, and state $\{q_m, 1, n\}$ are reachable, for a total of $m + 2^{n-2}$ states.

State $\{q_1\}$ is the initial state, and $\{q_i\}$ is reached by a^{i-1} for $i = 2, \ldots, m-1$. Also, $\{q_m, 1\}$ is reached by $a^{m-2}d$, and states q_m and 1 are present in every subset reachable from $\{q_m, 1\}$. By applying ab^{j-1} to $\{q_m, 1\}$ we reach $\{q_m, 1, j\}$; hence all subsets $\{q_m, 1\} \cup S$ with $|S| = 1$ are reachable. Assume now that we can reach all sets $\{q_m, 1\} \cup S$ with $|S| = k$, and suppose that we want to reach

$\{q_m, 1\} \cup T$ with $T = \{i_0, i_1, \ldots, i_k\}$ with $2 \leqslant i_0 < i_1 < \cdots < i_k \leqslant n-1$. This can be done by starting with $S = \{i_1 - i_0 + 1, \ldots, i_k - i_0 + 1\}$ and applying ab^{i_0-2}. Finally, to reach $\{q_m, 1, n\}$, start with $\{q_m, 1, n-1\}$ and apply d.

If $1 \leqslant i < j \leqslant m-1$, then state $\{q_i\}$ is distinguishable from $\{q_j\}$ by $a^{m-1-j}da^{n-1}d$. Also, state $i \in Q_n$ with $2 \leqslant i \leqslant n-1$ accepts $a^{n-1-i}d$ and no other state $j \in Q_n$ with $2 \leqslant j \leqslant n-1$ accepts this word. Hence, if $S, T \subseteq Q_n \setminus \{1, n\}$ and $S \neq T$, then $\{q_m, 1\} \cup S$ and $\{q_m, 1\} \cup T$ are distinguishable. State $\{q_k\}$ with $2 \leqslant k \leqslant m-1$ is distinguishable from state $\{q_m, 1\} \cup S$ because there is a word with a single d that is accepted from $\{q_m, 1\} \cup S$ but no such word is accepted by $\{q_k\}$. Hence all the non-final states are distinguishable, and $\{q_m, 1, n\}$ is the only final state. □

8 Conclusion

Our stream of right ideals is a universal witness for all common operations.

References

1. Bell, J., Brzozowski, J., Moreira, N., Reis, R.: Symmetric groups and quotient complexity of boolean operations. In: Esparza, J., Fraigniaud, P., Husfeldt, T., Koutsoupias, E. (eds.) ICALP 2014, Part II. LNCS, vol. 8573, pp. 1–12. Springer, Heidelberg (2014)
2. Brzozowski, J.: Quotient complexity of regular languages. J. Autom. Lang. Comb. 15(1/2), 71–89 (2010)
3. Brzozowski, J.: In search of the most complex regular languages. Internat. J. Found. Comput. Sci. 24(6), 691–708 (2013)
4. Brzozowski, J., Davies, G.: Most complex regular right-ideal languages (2013), `http://arxiv.org/abs/1311.4448`
5. Brzozowski, J., Jirásková, G., Li, B.: Quotient complexity of ideal languages. Theoret. Comput. Sci. 470, 36–52 (2013)
6. Brzozowski, J., Tamm, H.: Complexity of atoms of regular languages. Int. J. Found. Comput. Sci. 24(7), 1009–1027 (2013)
7. Brzozowski, J., Tamm, H.: Theory of átomata. Theoret. Comput. Sci. 539, 13–27 (2014)
8. Brzozowski, J., Ye, Y.: Syntactic complexity of ideal and closed languages. In: Mauri, G., Leporati, A. (eds.) DLT 2011. LNCS, vol. 6795, pp. 117–128. Springer, Heidelberg (2011)
9. Mirkin, B.G.: On dual automata. Kibernetika (Kiev) 2, 7–10 (1970) (Russian); English translation: Cybernetics 2, 6–9 (1966)
10. Piccard, S.: Sur les bases du groupe symétrique. Časopis Pro Pěstování Matematiky a Fysiky 68(1), 15–30 (1939)
11. Pin, J.E.: Syntactic semigroups. In: Handbook of Formal Languages. Word, Language, Grammar, vol. 1, pp. 679–746. Springer, New York (1997)
12. Yu, S.: State complexity of regular languages. J. Autom. Lang. Comb. 6, 221–234 (2001)