Theory of Átomata*

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. We show that every regular language defines a unique nondeterministic finite automaton (NFA), which we call "átomaton", whose states are the "atoms" of the language, that is, non-empty intersections of complemented or uncomplemented left quotients of the language. We describe methods of constructing the átomaton, and prove that it is isomorphic to the normal automaton of Sengoku, and to an automaton of Matz and Potthoff. We study "atomic" NFA's in which the right language of every state is a union of atoms. We generalize Brzozowski's double-reversal method for minimizing a deterministic finite automaton (DFA), showing that the result of applying the subset construction to an NFA is a minimal DFA if and only if the reverse of the NFA is atomic.

1 Introduction

Nondeterministic finite automata (NFA's) introduced by Rabin and Scott [9] in 1959 play a major role in the theory of automata. For many purposes it is necessary to convert an NFA to a deterministic finite automaton (DFA). In particular, for each NFA there exists a minimal DFA, unique up to isomorphism. This DFA is uniquely defined by every regular language, and uses the left quotients of the language as states. As well, it is possible to associate an NFA with each DFA, and this is the subject of the present paper. Our NFA is also uniquely defined by every regular language, and uses non-empty intersections of complemented and uncomplemented quotients—the "atoms" of the language—as states.

It appears that the NFA most often associated with a regular language is the universal automaton, sometimes appearing under different names. A recent substantial survey by Lombardy and Sakarovitch [7] on the subject of the universal automaton contains its history and a detailed discussion of its properties. We refer the reader to that paper, and mention only that research related to the universal automaton goes back to the 1970's: *e.g.*, in [3] as reported in [1], [4,6].

* This work was supported by the Natural Sciences and Engineering Research Council of Canada under grant No. OGP0000871, by the Estonian Center of Excellence in Computer Science, EXCS, financed by the European Regional Development Fund, and by the Estonian Science Foundation grant 7520.

G. Mauri and A. Leporati (Eds.): DLT 2011, LNCS 6795, pp. 105–116, 2011.

We call our NFA the "átomaton"[1] because it is based on the atoms of a regular language; we add the accent to minimize the possible confusion between "automaton" and "atomaton". Automata isomorphic to our átomaton have previously appeared in 1992 in the little-known master's thesis [10] of Sengoku, and in the 1995 paper [8] by Matz and Potthoff.

We introduce "atomic" automata, in which the right language of any state is a union of some atoms. This generalizes residual automata [5] in which the right language of any state is a left quotient (which we prove to be a union of atoms), and includes also átomata (where the right language of any state is an atom), trim DFA's, and the trim parts of universal automata.

We characterize the class of NFA's for which the subset construction yields a minimal DFA. More specifically, we show that the subset construction applied to a trim NFA produces a minimal DFA if and only if the reverse automaton of that NFA is atomic. This generalizes Brzozowski's method for DFA minimization by double reversal [2].

Section 2 recalls properties of regular languages, finite automata, and systems of language equations. Atoms of a regular language and the átomaton are introduced and studied in Section 3. In Section 4, we examine NFA's in which the right language of every state is a union of atoms. Brzozowski's method of DFA minimization is extended in Section 5, and Section 6 closes the paper.

2 Languages, Automata and Equations

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. The length of a word w is $|w|$. A *language* over Σ is any subset of Σ^*.

The following operations are defined on languages over Σ: *complement* ($\overline{L} = \Sigma^* \setminus L$), *union* ($K \cup L$), *intersection* ($K \cap L$), *product*, usually called *concatenation* or *catenation*, ($KL = \{w \in \Sigma^* \mid w = uv, u \in K, v \in L\}$), *positive closure* ($L^+ = \bigcup_{i \geqslant 1} L^i$), and *star* ($L^* = \bigcup_{i \geqslant 0} L^i$). The *reverse* w^R *of a word* $w \in \Sigma^*$ is defined as follows: $\varepsilon^R = \varepsilon$, and $(wa)^R = aw^R$. The *reverse of a language* L is denoted by L^R and defined as $L^R = \{w^R \mid w \in L\}$.

A *nondeterministic finite automaton* is a quintuple $\mathcal{N} = (Q, \Sigma, \delta, I, F)$, where Q is a finite, non-empty set of *states*, Σ is a finite non-empty *alphabet*, $\delta : Q \times \Sigma \to 2^Q$ 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 $\delta' : Q \times \Sigma^* \to 2^Q$, and $\delta'' : 2^Q \times \Sigma^* \to 2^Q$. We do not distinguish these functions notationally, but use δ for all three. The *language accepted* by an NFA $\mathcal{N}$ is $L(\mathcal{N}) = \{w \in \Sigma^* \mid \delta(I, w) \cap F \neq \emptyset\}$. Two NFA's are *equivalent* if they accept the same language. The *left language* of a state q of $\mathcal{N}$ is $L_{I,q}(\mathcal{N}) = \{w \in \Sigma^* \mid q \in \delta(I, w)\}$, and the *right language* of q is $L_{q,F}(\mathcal{N}) = \{w \in \Sigma^* \mid \delta(q, w) \cap F \neq \emptyset\}$. So $L(\mathcal{N}) = L_{I,F}(\mathcal{N})$. A state is *empty* if its right language is empty. Two states of an NFA are *equivalent* if their right languages are identical.

[1] The word should be pronounced with the accent on the first a.

A *deterministic finite automaton* 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. A DFA is an NFA in which the set of initial states is $\{q_0\}$ and the range of the transition function is restricted to singletons $\{q\}$, $q \in Q$.

An *incomplete deterministic finite automaton (IDFA)* is a quintuple $\mathcal{I} = (Q, \Sigma, \delta, q_0, F)$, where δ is a partial function such that either $\delta(q, a) = p$ for some $p \in Q$ or $\delta(q, a)$ is undefined. Every DFA is also an IDFA. An IDFA is *minimal* if no two of its states are equivalent.

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 well-known subset construction, where only subsets (including the empty subset) reachable from the initial subset of $\mathcal{N}^{\mathbb{D}}$ are used.

4. The *reversal* operation $\mathbb{R}$ applied to an NFA $\mathcal{N}$ yields an NFA $\mathcal{N}^{\mathbb{R}}$, where initial and final states of $\mathcal{N}$ are interchanged in $\mathcal{N}^{\mathbb{R}}$ and all the transitions between states are reversed.

2. The *trimming* operation $\mathbb{T}$ applied to an NFA $\mathcal{N}$ accepting a non-empty language deletes from $\mathcal{N}$ every state q not reachable from any initial state ($q \notin \delta(I, w)$ for any $w \in \Sigma^*$) and every state q that does not lead to any final state ($\delta(q, w) \cap F = \emptyset$ for all $w \in \Sigma^*$), along with the incident transitions. An NFA that has no such states is said to be *trim*. Note that, if $\mathcal{N}$ is trim, then so is $\mathcal{N}^{\mathbb{R}}$. Also, $\mathcal{N}^{\mathbb{R}\mathbb{T}} = \mathcal{N}^{\mathbb{T}\mathbb{R}}$ for any NFA $\mathcal{N}$.

If the trimming operation is applied to a DFA $\mathcal{D}$, we obtain the IDFA $\mathcal{D}^{\mathbb{T}}$, which behaves like $\mathcal{D}$, except that it does not have any empty states.

3. The *minimization* operation $\mathbb{M}$ applied to an IDFA (DFA) $\mathcal{D}$ yields the minimal IDFA (DFA) $\mathcal{D}^{\mathbb{M}}$ equivalent to $\mathcal{D}$.

A trim IDFA $\mathcal{I}$ is *bideterministic* if also $\mathcal{I}^{\mathbb{R}}$ is an IDFA. A language is *bideterministic* if its minimal IDFA is bideterministic.

Example 1. Figure 1 (a) shows an NFA $\mathcal{N}$. Its determinized DFA $\mathcal{N}^{\mathbb{D}}$ is in Fig. 1 (b), where braces around sets are omitted. The minimal equivalent $\mathcal{D} = \mathcal{N}^{\mathbb{D}\mathbb{M}}$ of $\mathcal{N}^{\mathbb{D}}$ is in Fig. 1 (c), where the equivalent states $\{2\}$, $\{1, 3\}$, and $\{2, 3\}$ are represented by $\{1, 3\}$. The reversed and trimmed version $\mathcal{D}^{\mathbb{R}\mathbb{T}}$ of the DFA $\mathcal{D}$ of Fig. 1 (c) is in Fig. 1 (d).

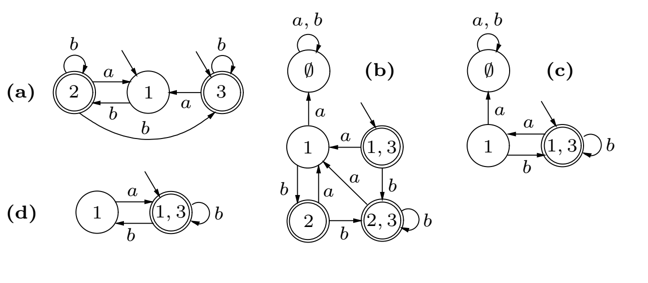

Fig. 1. (a) An NFA $\mathcal{N}$; (b) $\mathcal{N}^{\mathbb{D}}$; (c) $\mathcal{N}^{\mathbb{D}\mathbb{M}}$; (d) $\mathcal{N}^{\mathbb{D}\mathbb{M}\mathbb{R}\mathbb{T}}$

The *left quotient*, or simply *quotient,* of a language L by a word w is the language $w^{-1}L = \{x \in \Sigma^* \mid wx \in L\}$. Left quotients are also known as *right residuals*. Dually, the *right quotient* of a language L by a word w is the language $Lw^{-1} = \{x \in \Sigma^* \mid xw \in L\}$. Evidently, if $\mathcal{N}$ is an NFA and x is in $L_{I,q}(\mathcal{N})$, then $L_{q,F}(\mathcal{N}) \subseteq x^{-1}(L(\mathcal{N}))$.

The *quotient DFA* of a regular language L is $\mathcal{D} = (Q, \Sigma, \delta, q_0, F)$, where $Q = \{w^{-1}L \mid w \in \Sigma^*\}$, $\delta(w^{-1}L, a) = a^{-1}(w^{-1}L)$, $q_0 = \varepsilon^{-1}L = L$, and $F = \{w^{-1}L \mid \varepsilon \in w^{-1}L\}$. The *quotient IDFA* of a regular language L is $\mathcal{D}^{\mathbb{T}}$.

The following is from [7]: If $L \subseteq \Sigma^*$, a *subfactorization* of L is a pair (X, Y) of languages over Σ such that $XY \subseteq L$. A *factorization* of L is a subfactorisation (X, Y) such that, if $X \subseteq X'$, $Y \subseteq Y'$, and $X'Y' \subseteq L$ for any pair (X', Y'), then $X = X'$ and $Y = Y'$. The *universal automaton* of L is $\mathcal{U}_L = (Q, \Sigma, \delta, I, F)$ where Q is the set of all factorizations of L, $I = \{(X, Y) \in Q \mid \varepsilon \in X\}$, $F = \{(X, Y) \in Q \mid X \subseteq L\}$, and $(X', Y') \in \delta((X, Y), a)$ if and only if $Xa \subseteq X'$.

For any language L let $L^\varepsilon = \emptyset$ if $\varepsilon \notin L$ and $L^\varepsilon = \{\varepsilon\}$ otherwise. Also, let $n \geqslant 1$ and let $[n] = \{1, \ldots, n\}$. A *nondeterministic system of equations (NSE)* with n variables $L_1, \ldots, L_n$ is a set of language equations

$$L_i = \bigcup_{a \in \Sigma} a\Big(\bigcup_{j \in J_{i,a}} L_j\Big) \cup L_i^\varepsilon \quad i = 1, \ldots, n, \tag{1}$$

where $J_{i,a} \subseteq [n]$, together with an *initial set of variables* $\{L_i \mid i \in I\}$, where $I \subseteq [n]$ is an index set. The equations are assumed to have been simplified by the rules $a\emptyset = \emptyset$ and $K \cup \emptyset = \emptyset \cup K = K$, for any language K. Let $L_{i,a} = \bigcup_{j \in J_{i,a}} L_j$; then $L_{i,a} = a^{-1}L_i$ is the left quotient of L_i by a. The language defined by an NSE is $L = \bigcup_{i \in I} L_i$.

Each NSE defines a unique NFA $\mathcal{N}$ and *vice versa*. States of $\mathcal{N}$ correspond to the variables L_i, there is a transition $L_i \xrightarrow{a} L_j$ in $\mathcal{N}$ if and only if $j \in J_{i,a}$, the set of initial states of $\mathcal{N}$ is $\{L_i \mid i \in I\}$, and the set of final states is $\{L_i \mid L_i^\varepsilon = \{\varepsilon\}\}$.

If each L_i is a left quotient (that is, a right residual) of the language $L = \bigcup_{i \in I} L_i$, then the NSE and the corresponding NFA are called *residual* [5].

A *deterministic system of equations (DSE)* is an NSE

$$L_i = \bigcup_{a \in \Sigma} aL_{i_a} \cup L_i^\varepsilon \quad i = 1, \ldots, n, \tag{2}$$

where $i_a \in [n]$, $I = \{1\}$, and the empty language $\emptyset$ is retained if it appears.

Each DSE defines a unique DFA $\mathcal{D}$ and *vice versa*. Each state of $\mathcal{D}$ corresponds to a variable L_i, there is a transition $L_i \xrightarrow{a} L_j$ in $\mathcal{D}$ if and only if $i_a = j$, the initial state of $\mathcal{D}$ corresponds to L_1, and the set of final states is $\{L_i \mid L_i^\varepsilon = \{\varepsilon\}\}$. In the special case when $\mathcal{D}$ is minimal, its DSE constitutes its *quotient equations*, where every L_i is a quotient of the initial language L_1.

To simplify the notation, we write ε instead of $\{\varepsilon\}$ in equations.

Example 2. For the NFA of Fig. 1 (a), we have the NSE

$$\begin{aligned} L_1 &= bL_2, \\ L_2 &= aL_1 \cup b(L_2 \cup L_3) \cup \varepsilon, \\ L_3 &= aL_1 \cup bL_3 \cup \varepsilon, \end{aligned}$$

with the initial set $\{L_1, L_3\}$. The language $L = L_1 \cup L_3$ accepted by the DFA of Fig. 1 (b) is obtained from this NSE as shown by the equations below on the left. Renaming the unions of variables by new variables corresponding to subsets in the subset construction, we get the equations on the right; for example, $L_1 \cup L_3$ is renamed as $L_{\{1,3\}}$. This is the DSE for the DFA of Fig. 1 (b).

$$\begin{aligned} L_1 \cup L_3 &= aL_1 \cup b(L_2 \cup L_3) \cup \varepsilon, & L_{\{1,3\}} &= aL_{\{1\}} \cup bL_{\{2,3\}} \cup \varepsilon,\\ L_1 &= a\emptyset \cup bL_2, & L_{\{1\}} &= aL_{\emptyset} \cup bL_{\{2\}},\\ L_2 \cup L_3 &= aL_1 \cup b(L_2 \cup L_3) \cup \varepsilon, & L_{\{2,3\}} &= aL_{\{1\}} \cup bL_{\{2,3\}} \cup \varepsilon,\\ L_2 &= aL_1 \cup b(L_2 \cup L_3) \cup \varepsilon, & L_{\{2\}} &= aL_{\{1\}} \cup bL_{\{2,3\}} \cup \varepsilon,\\ \emptyset &= a\emptyset \cup b\emptyset. & L_{\emptyset} &= aL_{\emptyset} \cup bL_{\emptyset}. \end{aligned}$$

Noting that $L_{\{1,3\}}$, $L_{\{2,3\}}$, and $L_{\{2\}}$ are equivalent, we get the quotient equations for the DFA of Fig. 1 (c), where $L_{\{1\}} = a^{-1}L_{\{1,3\}}$, $L_{\{1,3\}} = b^{-1}L_{\{1,3\}}$, *etc.*

$$\begin{aligned} L_{\{1,3\}} &= aL_{\{1\}} \cup bL_{\{1,3\}} \cup \varepsilon,\\ L_{\{1\}} &= aL_{\emptyset} \cup bL_{\{1,3\}},\\ L_{\emptyset} &= aL_{\emptyset} \cup bL_{\emptyset}. \end{aligned}$$

3 The Átomaton of a Regular Language

From now on we consider only non-empty regular languages. Let L be a regular language, and let $L_1 = L, L_2, \ldots, L_n$ be its quotients. An *atom* of L is any non-empty language of the form $A = \widetilde{L_1} \cap \widetilde{L_2} \cap \cdots \cap \widetilde{L_n}$, where $\widetilde{L_i}$ is either L_i or $\overline{L_i}$, and at least one of the L_i is not complemented (in other words, $\overline{L_1} \cap \overline{L_2} \cap \cdots \cap \overline{L_n}$ is *not* an atom). A language has at most $2^n - 1$ atoms.

An atom is *initial* if it has L_1 (rather than $\overline{L_1}$) as a term; it is *final* if and only 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 $\widehat{L_1} \cap \widehat{L_1} \cap \cdots \cap \widehat{L_n}$, where $\widehat{L_i} = L_i$ if $\varepsilon \in L_i$, $\widehat{L_i} = \overline{L_i}$ otherwise. The atoms of L will be denoted by $A_1, \ldots, A_m$. By convention, I is the set of initial atoms and A_m is the final atom.

Proposition 1. *The following properties hold for atoms:*

1. *Atoms are pairwise disjoint, that is, $A_i \cap A_j = \emptyset$ for all $i, j \in [m]$, $i \neq j$.*
2. *The quotient $w^{-1}L$ of L by $w \in \Sigma^*$ is a (possibly empty) union of atoms.*
3. *The quotient $w^{-1}A_i$ of A_i by $w \in \Sigma^*$ is a (possibly empty) union of atoms.*

Proof. 1. If $A_i \neq A_j$, then there exists $h \in [n]$ such that L_h is a term of A_i and $\overline{L_h}$ is a term of A_j or *vice versa*. Hence $A_i \cap A_j = \emptyset$.

2. The empty quotient, if present, is the empty union of atoms. If $L_i \neq \emptyset$ is a quotient of L, then L_i is the union of all the 2^{n-1} intersections that have L_i as a term. This includes all the atoms that have L_i as a term, and possibly some empty intersections.

3. Consider the quotient equations of L. The quotient of each atom A_i by a letter $a \in \Sigma$ is an intersection X of complemented and uncomplemented quotients of L. If a quotient L_j of L does not appear as a term in X, then we "add

it in" by using the fact that $X = X \cap (L_j \cup \overline{L_j}) = (X \cap L_j) \cup (X \cap \overline{L_j})$. After all the missing quotients are so added, we obtain a union of atoms. Note that the intersection having all quotients complemented does not appear in this construction. It follows that $w^{-1}A_i$ is a union of atoms of L for every $w \in \Sigma^*$. □

Lemma 1. *Let $w, x \in \Sigma^*$. If $wx \in A_i$ and $x \in A_j$ then $wA_j \subseteq A_i$, for $i, j \in [m]$.*

Proof. Assume that $wx \in A_i$ and $x \in A_j$, but suppose $wy \notin A_i$ for some $y \in A_j$. Then $x \in w^{-1}A_i$ and $y \notin w^{-1}A_i$. By Proposition 1, Part 3, $w^{-1}A_i$ is a union of atoms. So, on the one hand, $x \in w^{-1}A_i$ and $x \in A_j$ together imply that $A_j \subseteq w^{-1}A_i$. On the other hand, from $y \notin w^{-1}A_i$ and $y \in A_j$, we get $A_j \not\subseteq w^{-1}A_i$. So if $wy \notin A_i$, we have a contradiction. Hence, $wA_j \subseteq A_i$. □

In the following definition we use a 1-1 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.

Definition 1. *Let $L = L_1 \subseteq \Sigma^*$ be any regular language with the set of atoms $Q = \{A_1, \ldots, A_m\}$, initial set of atoms $I \subseteq Q$, and final atom A_m. The* átomaton *of L is the NFA $\mathcal{A} = (\mathbf{Q}, \Sigma, \delta, \mathbf{I}, \{\mathbf{A}_m\})$, where $\mathbf{Q} = \{\mathbf{A}_i \mid A_i \in Q\}$, $\mathbf{I} = \{\mathbf{A}_i \mid A_i \in I\}$, and $\mathbf{A}_j \in \delta(\mathbf{A}_i, a)$ if and only if $aA_j \subseteq A_i$, for all $A_i, A_j \in Q$.*

Example 3. Let L be defined by the quotient equations below on the left and accepted by the quotient DFA of Fig. 2 (a).

$$\begin{aligned} L_1 &= aL_2 \cup bL_1, \\ L_2 &= aL_3 \cup bL_1 \cup \varepsilon, \\ L_3 &= aL_3 \cup bL_2. \end{aligned} \qquad \begin{aligned} L_{123} &= a(L_{123} \cup L_{\overline{1}23}) \cup b(L_{123} \cup L_{12\overline{3}}), \\ L_{\overline{1}23} &= aL_{\overline{1}\,\overline{2}3}, \\ L_{12\overline{3}} &= bL_{1\overline{2}\,\overline{3}}, \\ L_{\overline{1}\,\overline{2}3} &= b(L_{\overline{1}23} \cup L_{\overline{1}2\overline{3}}), \\ L_{1\overline{2}\,\overline{3}} &= a(L_{12\overline{3}} \cup L_{\overline{1}2\overline{3}}), \\ L_{\overline{1}2\overline{3}} &= \varepsilon. \end{aligned}$$

We find the atoms using the quotient equations. Thus

$$\begin{aligned} L_1 \cap L_2 \cap L_3 &= (aL_2 \cup bL_1) \cap (aL_3 \cup bL_1 \cup \varepsilon) \cap (aL_3 \cup bL_2) \\ &= (aL_2 \cap aL_3 \cap aL_3) \cup (bL_1 \cap bL_1 \cap bL_2) \\ &= a(L_2 \cap L_3) \cup b(L_1 \cap L_2) \\ &= a[(L_1 \cap L_2 \cap L_3) \cup (\overline{L_1} \cap L_2 \cap L_3)] \\ &\cup b[(L_1 \cap L_2 \cap L_3) \cup (L_1 \cap L_2 \cap \overline{L_3})], \textit{etc.} \end{aligned}$$

To simplify the notation, we denote $L_i \cap L_j$ by L_{ij}, $L_i \cap \overline{L_j}$ by $L_{i\overline{j}}$, *etc.* Noting that $L_{1\overline{2}3}$ is empty, we have the equations above on the right, from which we get Fig. 2 (b) for the átomaton of L.

Lemma 2. *For $w \in \Sigma^*$, $\mathbf{A}_j \in \delta(\mathbf{A}_i, w)$ if and only if $wA_j \subseteq A_i$, for $i, j \in [m]$.*

Proof. The proof is by induction on the length of w. If $|w| = 0$ and $\mathbf{A}_j \in \delta(\mathbf{A}_i, \varepsilon)$, then $i = j$ and $\varepsilon A_j \subseteq A_i$. If $|w| = 0$ and $\varepsilon A_j \subseteq A_i$, then $i = j$, since atoms are disjoint; hence $\mathbf{A}_j \in \delta(\mathbf{A}_i, \varepsilon)$. If $|w| = 1$, then the lemma holds by Definition 1.

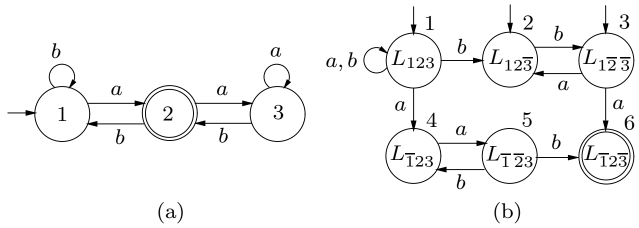

Fig. 2. (a) Quotient automaton; (b) átomaton of L

Now, let $w = av$, where $a \in \Sigma$ and $v \in \Sigma^+$, and assume that lemma holds for v. Suppose that $\mathbf{A}_j \in \delta(\mathbf{A}_i, av)$. Then there exists some state $\mathbf{A}_k$ such that $\mathbf{A}_k \in \delta(\mathbf{A}_i, a)$ and $\mathbf{A}_j \in \delta(\mathbf{A}_k, v)$. Thus, $aA_k \subseteq A_i$ by the definition of átomaton, and $vA_j \subseteq A_k$ by the induction assumption, implying that $avA_j \subseteq A_i$.

Conversely, let $avA_j \subseteq A_i$. Then $vA_j \subseteq a^{-1}A_i$. Let $x \in A_j$. Then $vx \in a^{-1}A_i$. Since by Proposition 1, Part 3, $a^{-1}A_i$ is a union of atoms, there exists some atom A_k such that $vx \in A_k$. Since $x \in A_j$, by Lemma 1 we get $vA_j \subseteq A_k$. Furthermore, because $avA_j \subseteq A_i$ and $x \in A_j$, we have $avx \in A_i$. Since $vx \in A_k$, then $aA_k \subseteq A_i$ by Lemma 1.

As the lemma holds for v and a, $vA_j \subseteq A_k$ implies $\mathbf{A}_j \in \delta(\mathbf{A}_k, v)$, and $aA_k \subseteq A_i$ implies $\mathbf{A}_k \in \delta(\mathbf{A}_i, a)$, showing that $\mathbf{A}_j \in \delta(\mathbf{A}_i, av)$. □

Proposition 2. *The right language of state $\mathbf{A}_i$ of átomaton $\mathcal{A}$ is the atom A_i, that is, $L_{\mathbf{A}_i,\{\mathbf{A}_m\}}(\mathcal{A}) = A_i$, for all $i \in [m]$.*

Proof. Let $w \in L_{\mathbf{A}_i,\{\mathbf{A}_m\}}(\mathcal{A})$; then $\mathbf{A}_m \in \delta(\mathbf{A}_i, w)$. By Lemma 2, we have $wA_m \subseteq A_i$. Since $\varepsilon \in A_m$, we have $w \in A_i$.

Now suppose that $w \in A_i$. Then $w\varepsilon \in A_i$, and since $\varepsilon \in A_m$, by Lemma 1 we get $wA_m \subseteq A_i$. By Lemma 2, $\mathbf{A}_m \in \delta(\mathbf{A}_i, w)$, that is, $w \in L_{\mathbf{A}_i,\{\mathbf{A}_m\}}(\mathcal{A})$. □

Theorem 1. *The language accepted by the átomaton $\mathcal{A}$ of L is L, that is, $L(\mathcal{A}) = L_{I,\{\mathbf{A}_m\}} = L$.*

Proof. We have $L(\mathcal{A}) = \bigcup_{A_i \in I} L_{\mathbf{A}_i,\{\mathbf{A}_m\}}(\mathcal{A}) = \bigcup_{A_i \in I} A_i$, by Proposition 2. Since I is the set of all atoms that have $L = L_1$ as a term, we also have $L = \bigcup_{A_i \in I} A_i$. □

Proposition 3. *The left language of state $\mathbf{A}_i$ of átomaton $\mathcal{A}$ of a language L is $L_{I,\mathbf{A}_i}(\mathcal{A}) = ((x^R)^{-1}L^R)^R$, for every $i \in [m]$ and every word x in A_i.*

Proof. If $w \in L_{I,\mathbf{A}_i}(\mathcal{A})$, then $\mathbf{A}_i \in \delta(\mathbf{A}_{i_0}, w)$ for some $\mathbf{A}_{i_0} \in \mathbf{I}$. Then $wA_i \subseteq A_{i_0}$ by Lemma 2. Since $A_{i_0} \subseteq L$, we also have $wA_i \subseteq L$, that is, $wx \in L$ for every $x \in A_i$. Then $x^R w^R \in L^R$, and $w^R \in (x^R)^{-1}L^R$. Thus, $w \in ((x^R)^{-1}L^R)^R$.

Now, let $w \in ((x^R)^{-1}L^R)^R$, where $x \in A_i$. Then $w^R \in (x^R)^{-1}L^R$, implying that $wx \in L$. By Theorem 1, there is some $\mathbf{A}_{i_0} \in I$ such that $wx \in L_{\mathbf{A}_{i_0},\{\mathbf{A}_m\}}(\mathcal{A})$. By Proposition 2, $wx \in A_{i_0}$. Since $x \in A_i$, by Lemma 1 we have $wA_i \subseteq A_{i_0}$. By Lemma 2, $\mathbf{A}_i \in \delta(\mathbf{A}_{i_0}, w)$, implying that $w \in L_{I,\mathbf{A}_i}(\mathcal{A})$. □

Proposition 4. *The left language of state* $\mathbf{A}_i$ *of átomaton* $\mathcal{A}$ *is non-empty, that is,* $L_{I,\mathbf{A}_i}(\mathcal{A}) \neq \emptyset$*, for every* $i \in [m]$.

Proof. Suppose that $L_{I,\mathbf{A}_i}(\mathcal{A}) = \emptyset$ for some $i \in [m]$. Then by Proposition 3, $((x^R)^{-1}L^R)^R = \emptyset$ for any $x \in A_i$. Then also $(x^R)^{-1}L^R = \emptyset$, implying that for any $w \in \Sigma^*$, $wx \notin L$. However, since there is some quotient L_j of L, $j \in [n]$, such that $A_i \subseteq L_j$, and there is an x in A_i, we have $x \in L_j$. Let $v \in \Sigma^*$ be such that $L_j = v^{-1}L$. Then we get $vx \in L$, which is a contradiction. □

Corollary 1. *The átomaton of any regular language is trim.*

Next we recall a (slightly modified version of a) theorem from [2]:

Theorem 2. *For a trim NFA* $\mathcal{N}$*,* $\mathcal{N}^{\mathbb{D}}$ *is minimal if* $\mathcal{N}^{\mathbb{R}}$ *is deterministic.*

We have defined a unique NFA, the átomaton, directly from the quotient equations of a language L, that is, from the minimal DFA recognizing L. In contrast to this, Sengoku [10] defined a unique NFA starting from any NFA accepting L: The *normal automaton* of L is the NFA $\mathcal{N}^{\mathbb{RDMTR}}$. Matz and Potthoff [8] (p. 78) defined an NFA $\overline{\mathcal{E}}$ as the reverse of the trim minimal DFA accepting L^R, that is $\overline{\mathcal{E}} = \mathcal{B}^{\mathbb{TR}}$, where $\mathcal{B}$ is the minimal DFA accepting L^R. We now relate a number of concepts associated with regular languages:

Theorem 3. *Let* L *be any regular language, and let* $\mathcal{A}$ *be its átomaton.*
1. *The reverse* $\mathcal{A}^{\mathbb{R}}$ *of* $\mathcal{A}$ *is an IDFA.*
2. $\mathcal{A}^{\mathbb{R}}$ *is minimal.*
3. *The determinization* $\mathcal{A}^{\mathbb{D}}$ *of* $\mathcal{A}$ *is the minimal DFA of* L.
4. *The normal NFA* $\mathcal{N}^{\mathbb{RDMTR}}$ *of any NFA accepting* L *is isomorphic to* $\mathcal{A}$.
5. *Matz and Potthoff's NFA* $\overline{\mathcal{E}}$ *is isomorphic to* $\mathcal{A}$.
6. $\mathcal{A}$ *is isomorphic to the quotient IDFA of* L *if and only if* L *is bideterministic.*

Proof. Suppose that L has the quotients $L_1, \ldots, L_n$ and atoms $A_1, \ldots, A_m$.

1. Since $\mathcal{A}$ has one accepting state, $\mathcal{A}^{\mathbb{R}}$ has one initial state. Because atoms are disjoint, a word w can belong to at most one atom. If w belongs to A_i, then, by Proposition 2, $w \in L_{\mathbf{A}_i,\{\mathbf{A}_m\}}(\mathcal{A})$ and $w \notin L_{\mathbf{A}_j,\{\mathbf{A}_m\}}(\mathcal{A})$ if $j \neq i$. Hence $w^R \in L_{\{\mathbf{A}_m\},\mathbf{A}_i}(\mathcal{A}^{\mathbb{R}})$ and $w^R \notin L_{\{\mathbf{A}_m\},\mathbf{A}_j}(\mathcal{A}^{\mathbb{R}})$ if $j \neq i$. Thus $\mathcal{A}^{\mathbb{R}}$ is an IDFA.

2. Since $\mathcal{A}$ is trim, so is $\mathcal{A}^{\mathbb{R}}$. Thus, if $\mathcal{A}^{\mathbb{R}}$ is not minimal, there must be states $\mathbf{A}_i, \mathbf{A}_j \in \mathbf{Q}$, $\mathbf{A}_i \neq \mathbf{A}_j$, such that $L_{\mathbf{A}_i,\mathbf{I}}(\mathcal{A}^{\mathbb{R}}) = L_{\mathbf{A}_j,\mathbf{I}}(\mathcal{A}^{\mathbb{R}})$. Let $L_k = u^{-1}L$ be any non-empty quotient of L, where $k \in [n]$ and $u \in \Sigma^*$. Then there are two possibilities: either $u \in L_{\mathbf{I},\mathbf{A}_i}(\mathcal{A})$, or $u \notin L_{\mathbf{I},\mathbf{A}_i}(\mathcal{A})$.

In the first case $u^R \in L_{\mathbf{A}_i,\mathbf{I}}(\mathcal{A}^{\mathbb{R}})$, and, since $L_{\mathbf{A}_i,\mathbf{I}}(\mathcal{A}^{\mathbb{R}}) = L_{\mathbf{A}_j,\mathbf{I}}(\mathcal{A}^{\mathbb{R}})$, we have $u^R \in L_{\mathbf{A}_j,\mathbf{I}}(\mathcal{A}^{\mathbb{R}})$, implying that $u \in L_{\mathbf{I},\mathbf{A}_j}(\mathcal{A})$. Thus, $L_{\mathbf{A}_i,\{\mathbf{A}_m\}}(\mathcal{A}) \subseteq L_k$ and $L_{\mathbf{A}_j,\{\mathbf{A}_m\}}(\mathcal{A}) \subseteq L_k$. In view of Proposition 2, A_i and A_j are both subsets of L_k.

Now, assume that $u \notin L_{\mathbf{I},\mathbf{A}_i}(\mathcal{A})$. Then, as in the first case, we get $u \notin L_{\mathbf{I},\mathbf{A}_j}(\mathcal{A})$. If $L_{\mathbf{A}_i,\{\mathbf{A}_m\}}(\mathcal{A}) \subseteq L_k$, then $ux \in L$ for some $x \in L_{\mathbf{A}_i,\{\mathbf{A}_m\}}(\mathcal{A})$. But then $x^R u^R \in L^R$. Since $x^R \in L_{\{\mathbf{A}_m\},\mathbf{A}_i}(\mathcal{A}^{\mathbb{R}})$ and $\mathcal{A}^{\mathbb{R}}$ is deterministic, we must have $u^R \in L_{\mathbf{A}_i,\mathbf{I}}(\mathcal{A}^{\mathbb{R}})$. This contradicts the assumption that $u \notin L_{\mathbf{I},\mathbf{A}_i}(\mathcal{A})$. Thus $L_{\mathbf{A}_i,\{\mathbf{A}_m\}}(\mathcal{A}) \not\subseteq L_k$, and similarly, $L_{\mathbf{A}_j,\{\mathbf{A}_m\}}(\mathcal{A}) \not\subseteq L_k$, implying that neither A_i nor A_j is a subset of L_k.

So, for every $k \in [n]$, either both atoms A_i and A_j are subsets of L_k or neither of them is. Since A_i and A_j are distinct, there must be an h such that $A_i \subseteq L_h$ and $A_j \subseteq \overline{L_h}$. This contradicts our earlier conclusion.

3. Since $\mathcal{A}^{\mathbb{R}}$ is deterministic, $\mathcal{A}^{\mathbb{D}}$ is minimal by Theorem 2.

4. Let $\mathcal{N}$ be any NFA accepting L. The DFA $\mathcal{N}^{\mathbb{RDM}}$ is the unique minimal DFA accepting the language L^R. By Parts 1 and 2, $\mathcal{A}^{\mathbb{R}}$ is a minimal IDFA, and it accepts $L^{\mathbb{R}}$. Since $\mathcal{N}^{\mathbb{RDMT}}$ is isomorphic to $\mathcal{A}^{\mathbb{R}}$, it follows that the normal automaton $\mathcal{N}^{\mathbb{RDMTR}}$ is isomorphic to $\mathcal{A}$.

5. Since $\mathcal{B}$ is isomorphic to $\mathcal{N}^{\mathbb{RDM}}$ of Part 4, the claim follows.

6. Let $\mathcal{D}$ be the quotient DFA of L, and suppose that $\mathcal{A}$ is isomorphic to $\mathcal{D}^{\mathbb{T}}$. By Part 1, $\mathcal{A}^R$ is an IDFA. Since $\mathcal{A}$ is isomorphic to $\mathcal{D}^{\mathbb{T}}$, $\mathcal{A}$ itself is an IDFA. Hence $\mathcal{A}$, and so also L, are bideterministic.

Conversely, let $\mathcal{B}$ be a trim bideterministic IDFA accepting L.

Since $\mathcal{B}^{\mathbb{R}}$ is deterministic, $\mathcal{B}^{\mathbb{D}}$ is minimal by Theorem 2. Since $\mathcal{B}$ is a trim IDFA, we have $\mathcal{B}^{\mathbb{DT}} = \mathcal{B}$; hence $\mathcal{B}$ is isomorphic to the quotient IDFA of L.

Since $\mathcal{B}$ is deterministic, $\mathcal{B}^{\mathbb{RD}}$ is minimal by Theorem 2, that is $\mathcal{B}^{\mathbb{RDM}} = \mathcal{B}^{\mathbb{RD}}$. Because $\mathcal{B}$ is trim, also $\mathcal{B}^{\mathbb{R}}$ is trim. Since $\mathcal{B}^{\mathbb{R}}$ is deterministic, we get $\mathcal{B}^{\mathbb{RDT}} = \mathcal{B}^{\mathbb{RT}} = \mathcal{B}^{\mathbb{R}}$. Thus $\mathcal{B}^{\mathbb{RDMTR}} = \mathcal{B}^{\mathbb{RDTR}} = \mathcal{B}^{\mathbb{RR}} = \mathcal{B}$ is the átomaton of L by Part 4. Hence $\mathcal{B}$ is both the minimal IDFA of L and its átomaton. □

As noted in [8], for each word w in L there is a unique path in $\mathcal{A}$ accepting w, and deleting any transition from $\mathcal{A}$ results in a smaller accepted language. It is also stated in [8] without proof that the right language $L_{q,F}(\mathcal{N})$ of any state q of an NFA $\mathcal{N}$ accepting L is a subset of a union of atoms. This holds because $L_{q,F}(\mathcal{N})$ is a subset of a (left) quotient of L, and quotients are unions of atoms by Proposition 1, Part 2.

Theorem 3 provides another method of finding the átomaton of L: simply trim the quotient DFA of L^R and reverse it. In view of this we have

Corollary 2. *If $L = L^R$ and L is accepted by quotient DFA $\mathcal{D}$, then $\mathcal{A} = \mathcal{D}^{\mathbb{T}}$.*

Example 4. Let $L = (b \cup ba)^*$; then $L^R = (b \cup ab)^*$, and it is accepted by the minimal DFA $\mathcal{D}$ of Fig. 1 (c). Its trimmed reverse is shown in Fig. 1 (d). Hence the NFA of Fig. 1 (d) is the átomaton of L.

4 Atomic Automata

We now introduce a new class of NFA's and study their properties.

Definition 2. *An NFA $\mathcal{N} = (Q, \Sigma, \delta, I, F)$ is* atomic *if for every state $q \in Q$, the right language $L_{q,F}(\mathcal{N})$ of q is a union of some atoms of $L(\mathcal{N})$.*

Note that, if $L_{q,F}(\mathcal{N}) = \emptyset$, then it is the union of zero atoms.

Recall that an NFA $\mathcal{N}$ is residual, if $L_{q,F}(\mathcal{N})$ is a (left) quotient of $L(\mathcal{N})$ for every $q \in Q$. Since every quotient is a union of atoms (see Proposition 1, Part 2), every residual NFA is atomic. However, the converse is not true: there exist atomic NFA's which are not residual. For example, the átomaton of Fig. 2

is atomic, but not residual. Note also that every trim DFA is a special case of a residual NFA; hence every trim DFA is atomic.

Let us now consider the universal automaton $\mathcal{U}_L = (Q, \Sigma, \delta, I, F)$ of a language L. We state some basic properties of this automaton based on [7]. Let (X, Y) be a factorization of L such that $X \neq \emptyset$ and $Y \neq \emptyset$. Then

(1) $Y = \bigcap_{x \in X} x^{-1}L$ and $X = \bigcap_{y \in Y} Ly^{-1}$.

(2) $L_{I,(X,Y)}(\mathcal{U}_L) = X$ and $L_{(X,Y),F}(\mathcal{U}_L) = Y$.

(3) The universal automaton $\mathcal{U}_L$ accepts L.

To recapitulate what was said above about residual NFA's and DFA's, and also to show that the trim part of the universal automaton is atomic, we have

Theorem 4. *Let L be any regular language. The following automata accepting L are atomic: 1. The átomaton $\mathcal{A}$. 2. Any trim DFA. 3. Any residual NFA. 4. The trim part of the universal automaton $\mathcal{U}_L$.*

Proof. 1. The right language of every state of $\mathcal{A}$ is an atom of L, so $\mathcal{A}$ is atomic.

2. The right language of every state of any trim DFA accepting L is a quotient of L. Since every quotient is a union of atoms, every trim DFA is atomic.

3. The right language of every state of any residual NFA of L is a quotient of L, and hence a union of atoms. Thus, any residual NFA is atomic.

4. Let (X, Y) be a state of $\mathcal{U}_L^{\mathbb{T}}$. Then $X, Y \neq \emptyset$. By (1) and (2), $L_{(X,Y),F}(\mathcal{U}_L) = \bigcap_{x \in X} x^{-1}L$. Let $L_1, \ldots, L_n$ be the quotients of L. Then for some $H \subseteq [n]$, $L_{(X,Y),F}(\mathcal{U}_L) = \bigcap_{i \in H} L_i$. Now $\bigcap_{i \in H} L_i = (\bigcap_{i \in H} L_i) \cap (\bigcap_{j \in [n] \setminus H} (L_j \cup \overline{L_j}) = \bigcup (\bigcap_{i \in H} L_i) \cap (\bigcap_{j \in [n] \setminus H} \widetilde{L_j})$, where $\widetilde{L_j}$ is either L_j or $\overline{L_j}$. Thus the right language of (X, Y) is a union of atoms of L. Since $L(\mathcal{U}_L^{\mathbb{T}}) = L(\mathcal{U}_L)$, and $L(\mathcal{U}_L) = L$ by (3), then $\mathcal{U}_L^{\mathbb{T}}$ is atomic. □

5 Extension of Brzozowski's Theorem on Minimal DFA's

Theorem 2 forms the basis for Brzozowski's "double-reversal" minimization algorithm [2]: Given any DFA (or IDFA) $\mathcal{D}$, reverse it to get $\mathcal{D}^{\mathbb{R}}$, determinize $\mathcal{D}^{\mathbb{R}}$ to get $\mathcal{D}^{\mathbb{R}\mathbb{D}}$, reverse $\mathcal{D}^{\mathbb{R}\mathbb{D}}$ to get $\mathcal{D}^{\mathbb{R}\mathbb{D}\mathbb{R}}$, and then determinize $\mathcal{D}^{\mathbb{R}\mathbb{D}\mathbb{R}}$ to get $\mathcal{D}^{\mathbb{R}\mathbb{D}\mathbb{R}\mathbb{D}}$. This last DFA is guaranteed to be minimal by Theorem 2, since $\mathcal{D}^{\mathbb{R}\mathbb{D}}$ is deterministic. Hence $\mathcal{D}^{\mathbb{R}\mathbb{D}\mathbb{R}\mathbb{D}}$ is the minimal DFA equivalent to $\mathcal{D}$.

Since this conceptually very simple algorithm carries out two determinizations, its complexity is exponential in the number of states of the original automaton in the worst case. However, its performance is good in practice, often better than Hopcrofts's algorithm [11,12]. Furthermore, this algorithm applied to an NFA still yields an equivalent minimal DFA; see [12], for example.

We now generalize Theorem 2:

Theorem 5. *For a trim NFA $\mathcal{N}$, $\mathcal{N}^{\mathbb{D}}$ is minimal if and only if $\mathcal{N}^{\mathbb{R}}$ is atomic.*

Proof. Let $\mathcal{N} = (Q, \Sigma, \delta, I, F)$ be any trim NFA, and let $\mathcal{N}^{\mathbb{R}} = (Q, \Sigma, \delta_{\mathbb{R}}, F, I)$ be its reverse. Let the atoms of $\mathcal{N}^{\mathbb{R}}$ be $B_1, \ldots, B_r$, and let $\mathcal{B}$ be the átomaton of $L(\mathcal{N}^{\mathbb{R}})$.

Assume first that $\mathcal{N}^{\mathbb{D}}$ is minimal. Let q be a state of $\mathcal{N}$, and hence of $\mathcal{N}^{\mathbb{R}}$; since $\mathcal{N}$ is trim, so is $\mathcal{N}^{\mathbb{R}}$, and there are $w, x \in \Sigma^*$ such that $x \in L_{F,q}(\mathcal{N}^{\mathbb{R}})$ and $w \in L_{q,I}(\mathcal{N}^{\mathbb{R}})$. Since $L_{q,I}(\mathcal{N}^{\mathbb{R}}) \subseteq x^{-1}L(\mathcal{N}^R)$, and every quotient of $L(\mathcal{N}^R)$ is a union of atoms, there is some $i \in [r]$ such that $w \in B_i$.

Suppose that $\mathcal{N}^{\mathbb{R}}$ is not atomic; then there must be a state $q' \in Q$ which is not a union of atoms. This means that there is some $i \in [r]$ and words $u, v \in B_i$ such that $u \in L_{q',I}(\mathcal{N}^{\mathbb{R}})$ but $v \notin L_{q',I}(\mathcal{N}^{\mathbb{R}})$. Suppose that z is in $L_{F,q'}(\mathcal{N}^{\mathbb{R}})$. Since $u \in z^{-1}L(\mathcal{N}^{\mathbb{R}})$, $u \in B_i$, and $z^{-1}L(\mathcal{N}^{\mathbb{R}})$ is a union of atoms, we must have $B_i \subseteq z^{-1}L(\mathcal{N}^{\mathbb{R}})$. But now $v \in B_i$ implies that $zv \in L(\mathcal{N}^{\mathbb{R}})$. Hence there must be a state $q'' \in Q$, $q'' \neq q'$, such that $v \in L_{q'',I}(\mathcal{N}^{\mathbb{R}})$. Therefore, we know that $u^R \in L_{I,q'}(\mathcal{N})$, $v^R \notin L_{I,q'}(\mathcal{N})$, and $v^R \in L_{I,q''}(\mathcal{N})$.

Now, since every state of $\mathcal{N}^{\mathbb{D}}$ is a subset of the state set Q of $\mathcal{N}$, there is a state s' of $\mathcal{N}^{\mathbb{D}}$ such that $q' \in s'$ and $u^R \in L_{I,s'}(\mathcal{N}^{\mathbb{D}})$, and there is a state s'' of $\mathcal{N}^{\mathbb{D}}$ such that $q'' \in s''$ and $v^R \in L_{I,s''}(\mathcal{N}^{\mathbb{D}})$. Since $v^R \notin L_{I,q'}(\mathcal{N})$, we have $q' \notin s''$, implying that $s' \neq s''$.

By Theorem 3, Part 4, $(\mathcal{N}^{\mathbb{R}})^{\mathbb{RDMTR}} = \mathcal{N}^{\mathbb{DMTR}}$ is isomorphic to the átomaton $\mathcal{B}$. By the assumption that $\mathcal{N}^{\mathbb{D}}$ is minimal, $\mathcal{N}^{\mathbb{DTR}}$ is isomorphic to $\mathcal{B}$. Thus $L_{s',I}(\mathcal{N}^{\mathbb{DTR}}) = B_k$ and $L_{s'',I}(\mathcal{N}^{\mathbb{DTR}}) = B_l$ for some $k, l \in [r]$. Since $u^R \in L_{I,s'}(\mathcal{N}^{\mathbb{D}})$, we have $u \in L_{s',I}(\mathcal{N}^{\mathbb{DR}}) = L_{s',I}(\mathcal{N}^{\mathbb{DRT}}) = L_{s',I}(\mathcal{N}^{\mathbb{DTR}}) = B_k$. This together with $u \in B_i$, yields $k = i$. Similarly, $v^R \in L_{I,s''}(\mathcal{N}^{\mathbb{D}})$ and $v \in B_i$, implies $l = i$. Thus, $L_{s',I}(\mathcal{N}^{\mathbb{DTR}}) = L_{s'',I}(\mathcal{N}^{\mathbb{DTR}})$. But then $L_{I,s'}(\mathcal{N}^{\mathbb{DT}}) = L_{I,s''}(\mathcal{N}^{\mathbb{DT}})$, which contradicts the inequality $s' \neq s''$. Therefore $\mathcal{N}^{\mathbb{R}}$ is atomic.

To prove the converse, assume that $\mathcal{N}^{\mathbb{R}}$ is atomic; then, for every state q of $\mathcal{N}^{\mathbb{R}}$, there is a set $H_q \subseteq [r]$ such that $L_{q,I}(\mathcal{N}^{\mathbb{R}}) = \bigcup_{i \in H_q} B_i$. This implies that $L_{I,q}(\mathcal{N}) = \bigcup_{i \in H_q} B_i^R$ for every state q of $\mathcal{N}$.

Let $\mathcal{N}^{\mathbb{D}} = (S, \Sigma, \gamma, I, G)$, and suppose that $\mathcal{N}^{\mathbb{D}}$ is not minimal. Then there are at least two states s' and s'' of $\mathcal{N}^{\mathbb{D}}$, $s' \neq s''$, with $L_{s',G}(\mathcal{N}^{\mathbb{D}}) = L_{s'',G}(\mathcal{N}^{\mathbb{D}})$. Let $\mathcal{D}_m = (Q_m, \Sigma, \delta_m, I, F_m)$ be a minimal DFA equivalent to $\mathcal{N}^{\mathbb{D}}$. Then there must be a state s of $\mathcal{D}_m$ such that $L_{s,F_m}(\mathcal{D}_m) = L_{s',G}(\mathcal{N}^{\mathbb{D}}) = L_{s'',G}(\mathcal{N}^{\mathbb{D}})$. Then $L_{I,s'}(\mathcal{N}^{\mathbb{D}}) \subset L_{I,s}(\mathcal{D}_m)$ and $L_{I,s''}(\mathcal{N}^{\mathbb{D}}) \subset L_{I,s}(\mathcal{D}_m)$ must also hold.

Since $\mathcal{N}^{\mathbb{DM}}$ is isomorphic to $\mathcal{D}_m$, and $\mathcal{N}^{\mathbb{DMTR}}$ is isomorphic to the átomaton $\mathcal{B}$ of $L(\mathcal{N}^{\mathbb{R}})$ by Part 4 of Theorem 3, also $(\mathcal{D}_m)^{\mathbb{TR}}$ is isomorphic to $\mathcal{B}$. Thus $L_{s,I}((\mathcal{D}_m)^{\mathbb{TR}}) = B_i$ for some $i \in [r]$. This implies that $L_{I,s}((\mathcal{D}_m)^{\mathbb{T}}) = B_i^R$. Thus $L_{I,s'}(\mathcal{N}^{\mathbb{D}}) \subset B_i^R$.

On the other hand, the left language of state s' of $\mathcal{N}^{\mathbb{D}}$ consists of all words u such that $u \in L_{I,q'}(\mathcal{N})$ for every $q' \in s'$, but $u \notin L_{I,q}(\mathcal{N})$ for any $q \notin s'$. That is, $L_{I,s'}(\mathcal{N}^{\mathbb{D}}) = (\bigcap_{q' \in s'} \bigcup_{i \in H_{q'}} B_i^R) \setminus (\bigcup_{q \notin s'} \bigcup_{i \in H_q} B_i^R)$. Since by Proposition 1, Part 1, $B_i \cap B_j = \emptyset$ for all $i, j \in [r]$, $i \neq j$, then also $B_i^R \cap B_j^R = \emptyset$, and the result of any boolean combination of sets B_i^R where $i \in [r]$, cannot be a proper subset of any B_i^R. Therefore, $L_{I,s'}(\mathcal{N}^{\mathbb{D}}) \subset B_i^R$ cannot hold and thus, $\mathcal{N}^{\mathbb{D}}$ must be minimal. □

Corollary 3. *If $\mathcal{N}$ is a non-minimal DFA, then $\mathcal{N}^{\mathbb{R}}$ is not atomic.*

Theorem 5 can be rephrased as follows: $\mathcal{N}$ is atomic if and only if $\mathcal{N}^{\mathbb{RD}}$ is minimal. Sengoku defines an NFA $\mathcal{N}$ to be *in standard form* [10] if and only if $\mathcal{N}^{\mathbb{RD}}$ is minimal, and also shows that the right language of every state of an NFA in standard form is equal to the union of right languages of some states of the normal automaton (that is, our átomaton).

6 Conclusions

We have introduced a natural set of languages—the atoms—that are defined by every regular language. We then defined a unique NFA, the átomaton, and related it to other known concepts. We introduced atomic automata, and generalized Brzozowski's method of minimization of DFA's by double reversal.

References

1. Arnold, A., Dicky, A., Nivat, M.: A note about minimal non-deterministic automata. Bull. EATCS 47, 166–169 (1992)
2. Brzozowski, J.A.: 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)
3. Carrez, C.: On the minimalization of non-deterministic automaton. Technical report, Lille University, Lille, France (1970)
4. Conway, J.: Regular Algebra and Finite Machines. Chapman and Hall, London (1971)
5. Denis, F., Lemay, A., Terlutte, A.: Residual finite state automata. Fund. Inform. 51, 339–368 (2002)
6. Kameda, T., Weiner, P.: On the state minimization of nondeterministic automata. IEEE Trans. Comput. C-19(7), 617–627 (1970)
7. Lombardy, S., Sakarovitch, J.: The universal automaton. In: Flum, J., Grädel, E., Wilke, T. (eds.) Logic and Automata: History and Perspectives. Texts in Logic and Games, vol. 2, pp. 457–504. Amsterdam University Press, Amsterdam (2007)
8. Matz, O., Potthoff, A.: Computing small finite nondeterministic automata. In: Engberg, U.H., Larsen, K.G., Skou, A. (eds.) Proceedings of the Workshop on Tools and Algorithms for Construction and Analysis of Systems. BRICS Note Series, pp. 74–88. BRICS, Aarhus (1995)
9. Rabin, M., Scott, D.: Finite automata and their decision problems. IBM J. Res. and Dev. 3, 114–129 (1959)
10. Sengoku, H.: Minimization of nondeterministic finite automata. Master's thesis, Kyoto University, Department of Information Science, Kyoto University, Kyoto, Japan (1992)
11. Tabakov, D., Vardi, M.: Experimental evaluation of classical automata constructions. In: Sutcliffe, G., Voronkov, A. (eds.) LPAR 2005. LNCS (LNAI), vol. 3835, pp. 396–411. Springer, Heidelberg (2005)
12. Watson, B.W.: Taxonomies and toolkits of regular language algorithms. PhD thesis, Faculty of Mathematics and Computing Science. Eindhoven University of Technology, Eindhoven, The Netherlands (1995)