Properties of Right One-Way Jumping Finite Automata

Simon Beier, Markus Holzer

Properties of Right One-Way Jumping Finite Automata

Simon Beier and Markus Holzer

Institut für Informatik, Universität Giessen,
Arndtstr. 2, 35392 Giessen, Germany
email: {simon.beier,holzer}@informatik.uni-giessen.de

Abstract. Right one-way jumping finite automata (ROWJFAs), were recently introduced in [H. Chigahara, S. Z. Fazekas, A. Yamamura: One-Way Jumping Finite Automata, *Internat. J. Found. Comput. Sci.*, 27(3), 2016] and are jumping automata that process the input in a discontinuous way with the restriction that the input head reads deterministically from left-to-right starting from the leftmost letter in the input and when it reaches the end of the input word, it returns to the beginning and continues the computation. We solve most of the open problems of these devices. In particular, we characterize the family of permutation closed languages accepted by ROWJFAs in terms of Myhill-Nerode equivalence classes. Using this, we investigate closure and non-closure properties as well as inclusion relations to other language families. We also give more characterizations of languages accepted by ROWJFAs for some interesting cases.

1 Introduction

Jumping finite automata [11] are a machine model for discontinuous information processing. Roughly speaking, a jumping finite automaton is an ordinary finite automaton, which is allowed to read letters from anywhere in the input string, not necessarily only from the left of the remaining input. In a series of papers [1, 6, 7, 13] different aspects of jumping finite automata were investigated, such as, e.g., inclusion relations, closure and non-closure results, decision problems, computational complexity of jumping finite automata problems, etc. Shortly after the introduction of jumping automata, a variant of this machine model was defined, namely (right) one-way jumping finite automata [3]. There the device moves the input head deterministically from left-to-right starting from the leftmost letter in the input and when it reaches the end of the input word, it returns to the beginning and continues the computation. As in the case of ordinary jumping finite automata inclusion relations to well-known formal language families, closure and non-closure results under standard formal language operations were investigated. Nevertheless, a series of problems on right one-way jumping automata (ROWJFAs) remained open in [3]. This is the starting point of our investigation.

First we develop a characterization of (permutation closed) languages that are accepted by ROWJFAs in terms of the Myhill-Nerode relation. It is shown that the permutation closed language L belongs to **ROWJ**, the family of all languages accepted by ROWJFAs, if and only if L can be written as the *finite union* of Myhill-Nerode equivalence classes. Observe, that the overall number of equivalence classes can be infinite. This result nicely contrasts the characterization of regular languages, which requires that the overall number of equivalence classes is finite. The characterization allows us to identify languages that are *not* accepted by ROWJFAs, which are useful to prove non-closure results on standard formal language operations. In this way we solve all of the open problems from [3] on the inclusion relations of ROWJFAs languages to other language families and on their closure properties. It is shown that the family **ROWJ** is an anti-abstract family of languages (anti-AFL), that is, it is not closed under any of the operations λ-free homomorphism, inverse homomorphism, intersection with regular sets, union, concatenation, or Kleene star. This is a little bit surprising for a language family defined by a deterministic automaton model. Although anti-AFLs are sometimes referred to an "unfortunate family of languages" there is linguistical evidence that such language families might be of crucial importance, since in [4] it was shown that the family of natural languages is an anti-AFL. On the other hand, the permutation closed languages in **ROWJ** almost form an anti-AFL, since this language family is closed under inverse homomorphism. Moreover, we obtain further characterizations of languages accepted by ROWJFAs. For instance, we show that

1. language wL is in **ROWJ** if and only if L is in **ROWJ**,
2. language Lw is in **ROWJ** if and only if L is regular, and
3. language L_1L_2 is in **ROWJ** if and only if L_1 is regular and L_2 is in **ROWJ**, where L_1 and L_2 have to fulfil some further easy pre-conditions.

The latter result is in similar vein as a result in [9] on linear context-free languages, where it was shown that L_1L_2 is a linear context-free language if and only if L_1 is regular and L_2 at most linear context free. Finally another characterization is given for letter bounded ROWJFA languages, namely, the language $L \subseteq a_1^*a_2^* \dots a_n^*$ is in **ROWJ** if and only if L is regular. This result nicely generalizes the fact that every unary language accepted by an ROWJFA is regular.

2 Preliminaries

We assume the reader to be familiar with the basics in automata and formal language theory as contained, for example, in [10]. Let $\mathbb{N} = \{0, 1, 2, \dots\}$ be the set of non-negative integers. We use $\subseteq$ for inclusion, and $\subset$ for proper inclusion. Let Σ be an alphabet. Then Σ^* is the set of all words over Σ, including the empty word λ. For a language $L \subseteq \Sigma^*$ define the set $\mathsf{perm}(L) = \cup_{w\in L} \mathsf{perm}(w)$, where $\mathsf{perm}(w) = \{\, v \in \Sigma^* \mid v \text{ is a permutation of } w \,\}$. Then a language L is called *permutation closed* if $L = \mathsf{perm}(L)$. The length of a word $w \in \Sigma^*$ is denoted by $|w|$. For the number of occurrences of a symbol a in w we use the notation $|w|_a$. We denote the powerset of a set S by 2^S. For $\Sigma = \{a_1, a_2, \ldots, a_k\}$, the *Parikh-mapping* $\psi : \Sigma^* \to \mathbb{N}^k$ is the function $w \mapsto (|w|_{a_1}, |w|_{a_2}, \ldots, |w|_{a_k})$. A language $L \subseteq \Sigma^*$ is called *semilinear* if its *Parikh-image* $\psi(L)$ is a semilinear subset of $\mathbb{N}^k$, a definition of those can be found in [8].

The elements of $\mathbb{N}^k$ can be partially ordered by the $\leq$-relation on vectors. For vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{N}^k$ we write $\boldsymbol{x} \leq \boldsymbol{y}$ if all components of $\boldsymbol{x}$ are less or equal to the corresponding components of $\boldsymbol{y}$. The value $||\boldsymbol{x}||$ is the maximum norm of $\boldsymbol{x}$, that is, $||(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k)|| = \max\{\, |\boldsymbol{x}_i| \mid 1 \leq i \leq k \,\}$.

For $v, w \in \Sigma^*$, we say that v is a prefix of w if there is an $x \in \Sigma^*$ with $w = vx$. Moreover, v is a sub-word of w if there are $x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_{n+1} \in \Sigma^*$ with $v = x_1 x_2 \cdots x_n$ and $w = y_1 x_1 y_2 x_2 \cdots y_n x_n y_{n+1}$, for some $n \geq 0$. A language $L \subseteq \Sigma^*$ is called prefix-free if and only if there are no words $v, w \in L$ such that $v \neq w$ and v is a prefix of w.

For an alphabet Σ and a language $L \subseteq \Sigma^*$, let $\sim_L$ be the *Myhill-Nerode equivalence relation* on Σ^*. So, for $v, w \in \Sigma^*$, we have $v \sim_L w$ if and only if, for all $u \in \Sigma^*$, the equivalence $vu \in L \Leftrightarrow wu \in L$ holds. For $w \in \Sigma^*$, we call the equivalence class $[w]_{\sim_L}$ positive if and only if $w \in L$. Otherwise, the equivalence class $[w]_{\sim_L}$ is called negative.

A *deterministic finite automaton*, a DFA for short, is defined as a tuple $A = (Q, \Sigma, R, s, F)$, where Q is the finite set of states, Σ is the finite input alphabet, $\Sigma \cap Q = \emptyset$, R is a partial function from $Q \times \Sigma$ to Q, $s \in Q$ is the start state, and $F \subseteq Q$ is the set of final states. The elements of R are referred to a rules of A and we write $py \to q \in R$ instead of $R(p, y) = q$. A configuration of A is a string in $Q\Sigma^*$. A DFA makes a transition from configuration paw to configuration qw if $pa \to q \in R$, where $p, q \in Q$, $a \in \Sigma$, and $w \in \Sigma^*$. We denote this by $paw \vdash_A qw$ or just $paw \vdash qw$ if it is clear which DFA we are referring to. In the standard manner, we extend $\vdash$ to $\vdash^n$, where $n \geq 0$. Let $\vdash^+$ and $\vdash^*$ denote the transitive closure of $\vdash$ and the transitive-reflexive closure of $\vdash$, respectively. Then, the language accepted by A is $L(A) = \{\, w \in \Sigma^* \mid \exists f \in F : sw \vdash^* f \,\}$. We say that A accepts $w \in \Sigma^*$ if $w \in L(A)$ and that A rejects w otherwise. The family of languages accepted by DFAs is referred to as **REG**.

A *jumping finite automaton*, a JFA for short, is a tuple $A = (Q, \Sigma, R, s, F)$, where Q, Σ, R, s, and F are the same as in the case of DFAs. A configuration of A is a string in $\Sigma^* Q \Sigma^*$. The binary jumping relation, symbolically denoted by $\curvearrowright_A$, over $\Sigma^* Q \Sigma^*$ is defined as follows. Let x, z, x', z' be strings in Σ^* such that $xz = x'z'$ and $py \to q \in R$. Then, the automaton A makes a jump from $xpyz$ to $x'qz'$, symbolically written as $xpyz \curvearrowright_A x'qz'$ or just $xpyz \curvearrowright x'qz'$ if it is clear which JFA we are referring to. In the standard manner, we extend $\curvearrowright$ to $\curvearrowright^n$, where $n \geq 0$. Let $\curvearrowright^+$ and $\curvearrowright^*$ denote the transitive closure of $\curvearrowright$ and the transitive-reflexive closure of $\curvearrowright$, respectively. Then, the language accepted by A is $L(A) = \{\, uv \mid u, v \in \Sigma^*, \exists f \in F : usv \curvearrowright^* f \,\}$. We say that A accepts $w \in \Sigma^*$ if $w \in L(A)$ and that A rejects w otherwise. Let **JFA** be the family of all languages that are accepted by JFAs.

A *right one-way jumping finite automaton*, a ROWJFA for short, is a tuple $A = (Q, \Sigma, R, s, F)$, where Q, Σ, R, s, and F are defined as in a DFA. A configuration of A is a string in $Q\Sigma^*$. The right one-way jumping relation, symbolically denoted by $\circlearrowright_A$, over $Q\Sigma^*$ is defined as follows. For $p \in Q$ we set

$$\Sigma_p = \Sigma_{R,p} = \{\, b \in \Sigma \mid pb \to q \in R \text{ for some } q \in Q \,\}\,.$$

Now, let $pa \to q \in R$, $x \in (\Sigma \setminus \Sigma_p)^*$, and $y \in \Sigma^*$. Then, the ROWJFA A makes a jump from the configuration $pxay$ to the configuration qyx, symbolically written as $pxay \circlearrowright_A qyx$ or just $pxay \circlearrowright qyx$ if it is clear which ROWJFA we are referring to. In the standard manner, we extend $\circlearrowright$ to $\circlearrowright^n$, where $n \geq 0$. Let $\circlearrowright^+$ and $\circlearrowright^*$ denote the transitive closure of $\circlearrowright$ and the transitive-reflexive closure of $\circlearrowright$, respectively. The language accepted by A is $L(A) = \{\, w \in \Sigma^* \mid \exists f \in F : sw \circlearrowright^* f \,\}$. We say that A accepts $w \in \Sigma^*$ if $w \in L(A)$ and that A rejects w otherwise. Let **ROWJ** be the family of all languages that are accepted by ROWJFAs. Furthermore, for $n \geq 0$, let $\mathbf{ROWJ_n}$ be the class of all languages accepted by ROWJFAs with at most n accepting states.

Besides the above mentioned language families let **FIN**, **DCF**, **CF**, and **CS** be the families of finite, deterministic context-free, context-free, and context-sensitive languages. Moreover, we are interested in permutation closed language families. These language families are referred to by a prefix **p**. E.,g., **pROWJ** denotes the language family of all permutation closed **ROWJ** languages.

Sometimes, for a DFA A, we will also consider the relations $\curvearrowright$ and $\circlearrowright$, that we get by interpreting A as a JFA or a ROWJFA. The following three languages are associated to A:

- $L_D(A)$ is the language accepted by A, interpreted as an ordinary DFA.
- $L_J(A)$ is the language accepted by A, interpreted as an JFA.
- $L_R(A)$ is the language accepted by A, interpreted as an ROWJFA.

From a result in [12] and from [3, Theorem 10], we get

$$L_D(A) \subseteq L_R(A) \subseteq L_J(A) = \mathsf{perm}(L_D(A)). \tag{1}$$

As a consequence, we have **JFA** = **pJFA**. Next we give an example of a DFA A with $L_D(A) \subset L_R(A) \subset L_J(A)$.

Example 1. Let A be the DFA

$$A = (\{q_0, q_1, q_2, q_3\}, \{a, b\}, R, q_0, \{q_3\})\,,$$

where R consists of the rules $q_0b \to q_1$, $q_0a \to q_2$, $q_2b \to q_3$, and $q_3a \to q_2$. The automaton A is depicted in Figure 1.

It holds $L_D(A) = (ab)^+$ and

$$L_J(A) = \mathsf{perm}\left((ab)^+\right) = \left\{\, w \in \{a,b\}^+ \;\middle|\; |w|_a = |w|_b \,\right\}.$$

To show how ROWJFAs work, we give an example computation of A, interpreted as an ROWJFA, on the input word $aabbba$:

$$q_0aabbba \circlearrowright q_2abbba \circlearrowright q_3bbaa \circlearrowright q_2abb \circlearrowright q_3ba \circlearrowright q_2b \circlearrowright q_3$$

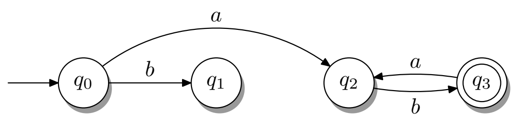

Fig. 1. The automaton A with $L_D(A) \subset L_R(A) \subset L_J(A)$.

That shows $aabbba \in L_R(A)$. Analogously, one can see that every word that contains the same number of a's and b's and that begins with an a is in $L_R(A)$. On the other hand, no other word can be accepted by A, interpreted as an ROWJFA. So, we get $L_R(A) = \{\, w \in a\{a,b\}^* \mid |w|_a = |w|_b \,\}$. Notice that this language is non-regular and not closed under permutation. $\square$

The following basic property will be used later on.

Lemma 2. *Let $A = (Q, \Sigma, R, s, F)$ be a DFA. Consider two words $v, w \in \Sigma^*$, states $p, q \in Q$, and an $n \geq 0$ with $pv \circlearrowright^n qw$. Then, there is a word $x \in \Sigma^*$ such that xw is a permutation of v, and $px \vdash^n q$.*

Proof. We prove this by induction on n. If $n = 0$, we have $pv = qw$ and just set $x = \lambda$. Now, assume $n > 0$ and that the lemma is true for the relation $\circlearrowright^{n-1}$. We get a state $r \in Q$, a symbol $a \in \Sigma_r$, and words $y \in (\Sigma \setminus \Sigma_r)^*$ and $z \in \Sigma^*$ such that $w = zy$ and $pv \circlearrowright^{n-1} ryaz \circlearrowright qw$. By the induction hypothesis, there is an $x' \in \Sigma^*$ such that $x'yaz$ is a permutation of v, and $px' \vdash^{n-1} r$. Set $x = x'a$. Then, the word $xw = x'azy$ is a permutation of $x'yaz$, which is a permutation of v. Furthermore, we get $px = px'a \vdash^{n-1} ra \vdash q$. This proves the lemma. $\square$

3 A Characterization of Permutation Closed Languages Accepted by ROWJFAs

By the Myhill-Nerode theorem, a language L is regular if and only if the Myhill-Nerode relation $\sim_L$ has only a finite number of equivalence classes. Moreover, the number of equivalence classes equals the number of states of the minimal DFA accepting L, see for example [10]. We can give a similar characterization for permutation closed languages that are accepted by an ROWJFA.

Theorem 3. *Let L be a permutation closed language and $n \geq 0$. Then, the language L is in $\mathbf{ROWJ_n}$ if and only if the Myhill-Nerode relation $\sim_L$ has at most n positive equivalence classes.*

Proof. First, assume that L is in $\mathbf{ROWJ_n}$ and let $A = (Q, \Sigma, R, s, F)$ be a DFA with $|F| \leq n$ and $L_R(A) = L$. Consider $v, w \in L$ and $f \in F$ with $sv \circlearrowright^* f$ and $sw \circlearrowright^* f$. Lemma 2 shows that there are permutations v' and w' of v and w with $sv' \vdash^* f$ and $sw' \vdash^* f$. Because language L is closed under permutation we have $v \sim_L v'$ and $w \sim_L w'$. Now, let $u \in \Sigma^*$. Thus $sv'u \circlearrowright^* fu$ and $sw'u \circlearrowright^* fu$. That gives us

$$v'u \in L \Leftrightarrow (\exists g \in F : fu \circlearrowright^* g) \Leftrightarrow w'u \in L.$$

We have shown $v \sim_L v' \sim_L w' \sim_L w$. From $L = \bigcup_{f \in F} \{ w \in \Sigma^* \mid sw \circlearrowright^* f \}$, we get $|L/\sim_L| \leq |F| \leq n$, which means that $\sim_L$ has at most n positive equivalence classes.

Assume now that $\sim_L$ has at most n positive equivalence classes and let $\Sigma = \{a_1, a_2, \ldots, a_k\}$ be an alphabet with $L \subseteq \Sigma^*$. Set $L_\lambda = L \cup \{\lambda\}$. Define the map $S : L_\lambda/\sim_L \to 2^{\mathbb{N}^k}$ through $[w] \mapsto \{ \boldsymbol{x} \in \mathbb{N}^k \setminus \boldsymbol{0} \mid \psi^{-1}(\psi(w) + \boldsymbol{x}) \subseteq L \}$. The definition of $\sim_L$ and the fact that L is closed under permutation make the map S well-defined. Consider the relation $\leq$ on $\mathbb{N}^k$. For each $[w] \in L_\lambda/\sim_L$, let $M([w])$ be the set of minimal elements of $S([w])$. So, for every $[w] \in L_\lambda/\sim_L$ and $\boldsymbol{x} \in S([w])$, there is an $\boldsymbol{x_0} \in M([w])$ such that $\boldsymbol{x_0} \leq \boldsymbol{x}$. Due to [5] each subset of $\mathbb{N}^k$ has only a finite number of minimal elements, so the sets $M([w])$ are finite. For $i \in \{1, 2, \ldots, k\}$, let $\pi_i : \mathbb{N}^k \to \mathbb{N}$ be the canonical projection on the ith factor and set

$$m_i = \max \left(\bigcup_{[w] \in L_\lambda/\sim_L} \{ \pi_i(\boldsymbol{x}) \mid \boldsymbol{x} \in M([w]) \} \right),$$

where $\max(\emptyset)$ should be 0. We have $m_i < \infty$, for all $i \in \{1, 2, \ldots, k\}$, because of $|L_\lambda/\sim_L| \leq n + 1$. Let

$$Q = \left\{ q_{[wv]_{\sim_L}} \;\middle|\; w \in L_\lambda,\ v \in \Sigma^* \text{ with } |v|_{a_i} \leq m_i, \text{ for all } i \in \{1, 2, \ldots, k\} \right\}$$

be a set of states. The finiteness of $L_\lambda/\sim_L$ implies that Q is also finite. Set

$$F = \left\{ q_{[w]_{\sim_L}} \;\middle|\; w \in L \right\} \subseteq Q.$$

We get $|F| = |L/\sim_L| \leq n$. Define the partial mapping $R : Q \times \Sigma \to Q$ by $R(q_{[y]_{\sim_L}}, a) = q_{[ya]_{\sim_L}}$, if $q_{[ya]_{\sim_L}} \in Q$, and $R(q_{[y]_{\sim_L}}, a)$ be undefined otherwise, for $a \in \Sigma$ and $y \in \Sigma^*$ with $q_{[y]_{\sim_L}} \in Q$. Consider the DFA $A = (Q, \Sigma, R, q_{[\lambda]_{\sim_L}}, F)$. We will show that $L_R(A) = L$.

First, let $y \in L_R(A)$. Then, there exists $w \in L$ with $q_{[\lambda]_{\sim_L}} y \circlearrowright^* q_{[w]_{\sim_L}}$. From Lemma 2 it follows that there is a permutation y' of y with $q_{[\lambda]_{\sim_L}} y' \vdash^* q_{[w]_{\sim_L}}$. Now, the definition of R tells us $y' \sim_L w$. We get $y' \in L$ and also $y \in L$, because L is closed under permutation. That shows $L_R(A) \subseteq L$.

Now, let $y \in \Sigma^* \setminus L_R(A)$. There are two possibilities:

1. There is a $w \in \Sigma^* \setminus L$ with $q_{[w]_{\sim_L}} \in Q$ such that $q_{[\lambda]_{\sim_L}} y \circlearrowright^* q_{[w]_{\sim_L}}$. Then, there is a permutation y' of y with $q_{[\lambda]_{\sim_L}} y' \vdash^* q_{[w]_{\sim_L}}$. We get $y' \sim_L w$. It follows $y' \notin L$, which gives us $y \notin L$.
2. There is a $w \in L_\lambda$, a $v \in \Sigma^*$ with $|v|_{a_i} \leq m_i$, for all $i \in \{1, 2, \ldots, k\}$, and a $z \in (\Sigma \setminus \Sigma_{q_{[wv]_{\sim_L}}})^+$ such that $q_{[\lambda]_{\sim_L}} y \circlearrowright^* q_{[wv]_{\sim_L}} z$. By Lemma 2 there is a $y' \in \Sigma^*$ such that $y'z$ is a permutation of y and $q_{[\lambda]_{\sim_L}} y' \vdash^* q_{[wv]_{\sim_L}}$. We get $y' \sim_L wv$. Set

$$U = \bigcup_{t \in \Sigma^*} \{ u \in \Sigma^* \mid ut \in \mathsf{perm}(v) \text{ and } wu \in L_\lambda \}.$$

We have $\lambda \in U$. Let $u_0 \in U$ such that $|u_0| = \max(\{\, |u| \mid u \in U \,\})$ and let $t_0 \in \Sigma^*$ such that $u_0 t_0 \in \mathsf{perm}(v)$. It follows that $|t_0|_{a_i} \leq |v|_{a_i} \leq m_i$, for all $i \in \{1, 2, \ldots, k\}$, and that there exists *no* $\boldsymbol{x} \in M([wu_0]_{\sim_L})$ with $\boldsymbol{x} \leq \psi(t_0)$. Otherwise, we would have an $x' \in \psi^{-1}(\boldsymbol{x})$ which is a non-empty sub-word of t_0 such that $wu_0x' \in L$, which implies $u_0x' \in U$. However, this is a contradiction to the maximality of $|u_0|$. That shows that there is no $\boldsymbol{x} \in M([wu_0]_{\sim_L})$ with $\boldsymbol{x} \leq \psi(t_0)$. Let now $\boldsymbol{x_0} \in M([wu_0]_{\sim_L})$. There exists a $j \in \{1, 2, \ldots, k\}$ with $|t_0|_{a_j} < \pi_j(\boldsymbol{x_0}) \leq m_j$. Because of $|t_0|_{a_i} \leq m_i$, for all i with $i \in \{1, 2, \ldots, k\}$, and $z \in (\Sigma \setminus \Sigma_{q_{[wv]_{\sim_L}}})^+ = (\Sigma \setminus \Sigma_{q_{[wu_0t_0]_{\sim_L}}})^+$, we get $|z|_{a_j} = 0$. That gives $|t_0 z|_{a_j} < \pi_j(\boldsymbol{x_0})$ and that $\psi(t_0 z) \geq \boldsymbol{x_0}$ is false. So, we have shown $\psi(t_0 z) \notin S([wu_0]_{\sim_L})$, which implies $wu_0t_0z \notin L$. From $wu_0t_0z \sim_L wvz \sim_L y'z \sim_L y$, it follows that $y \notin L$.

We have seen $L_R(A) = L$. This shows that L is in $\mathbf{ROWJ_n}$. □

The previous theorem allows us to determine for a lot of interesting languages whether they belong to **ROWJ** or not.

Corollary 4. *Let L be a permutation closed language. Then, the language L is in* **ROWJ** *if and only if the Myhill-Nerode relation $\sim_L$ has only a finite number of positive equivalence classes.* □

An application of the last corollary is the following.

Lemma 5. *The language $L = \{\, w \in \{a, b\}^* \mid |w|_b = 0 \vee |w|_b = |w|_a \,\}$ is not included in* **ROWJ**.

Proof. The language L is closed under permutation. For $\sim_L$, the positive equivalence classes $[a^0], [a^1], \ldots$ are pairwise different, since $a^n b^m \in L$ if and only if $m \in \{0, n\}$. Corollary 4 tells us that L is not in **ROWJ**. □

There are counterexamples for both implications of Corollary 4, if we do not assume that the language L is closed under permutation. For instance, set $L = \{\, a^n b^n \mid n \geq 0 \,\}$, which was shown to be not in **ROWJ** in [3]. Then, the positive equivalence classes of $\sim_L$ are $[\lambda]$ and $[ab]$. On the other hand, we have:

Lemma 6. *There is a language L in* **ROWJ** *such that $\sim_L$ has an infinite number of positive equivalence classes.* □

From Corollary 4 we conclude the following equivalence.

Corollary 7. *Let L be a permutation closed* **ROWJ** *language over the alphabet Σ. Then, the language L is regular if and only if $\Sigma^* \setminus L$ is in* **ROWJ**. □

The previous corollary gives us:

Lemma 8. *The language $\{\, w \in \{a, b\}^* \mid |w|_a \neq |w|_b \,\}$ is not in* **ROWJ**. □

Having the statement of Theorem 3, it is natural to ask, which numbers arise as the number of positive equivalence classes of the Myhill-Nerode relation $\sim_L$ of a permutation closed language L. The answer is, that all natural numbers arise this way, even if we restrict ourselves to some special families:

Theorem 9. *For each $n > 0$, there is a permutation closed language which is (1) finite, (2) regular, but infinite, (3) context-free, but non-regular, (4) non-context-free such that the corresponding Myhill-Nerode relation has exactly n positive equivalence classes.* □

The previous theorem, together with Theorem 3, implies that the language families $\mathbf{ROWJ_n}$ form a proper hierarchy, even if we only consider languages out of special language families:

Corollary 10. *For all $n \geq 0$, we have $\mathbf{ROWJ_n} \cap \mathbf{FAM} \subset \mathbf{ROWJ_{n+1}} \cap \mathbf{FAM}$, where $\mathbf{FAM}$ is either 2^{Σ^*}, $\mathbf{FIN}$, $\mathbf{REG} \setminus \mathbf{FIN}$, $\mathbf{CF} \setminus \mathbf{REG}$, or $\mathbf{CS} \setminus \mathbf{CF}$. The statement remains valid if restricted to permutation closed languages.* □

4 Inclusion Relations Between Language Families

We investigate inclusion relations between **ROWJ** and other important languages families. The following relations were given in [3]: (1) $\mathbf{REG} \subset \mathbf{ROWJ}$, (2) **ROWJ** and **CF** are incomparable, and (3) $\mathbf{ROWJ} \nsubseteq \mathbf{JFA}$. It was stated as an open problem if $\mathbf{JFA} \subset \mathbf{ROWJ}$. We can answer this using Lemma 5:

Theorem 11. *The language families* **ROWJ** *and* **JFA** *are incomparable.* □

For the complexity of **ROWJ**, we get that the language family **ROWJ** is included in both of the complexity classes $\mathbf{DTIME}(n^2)$ and $\mathbf{DSPACE}(n)$. This implies that **ROWJ** is properly included in **CS**. Moreover, we find the following relations:

Theorem 12. *We have (1) $\mathbf{ROWJ} \subset \mathbf{CS}$, (2)* **ROWJ** *and* **DCF** *are incomparable, and (3) every language in* **ROWJ** *is semilinear.* □

For permutation closed language families the next theorem applies.

Theorem 13. *We have $\mathbf{pFIN} \subset \mathbf{pREG} \subset \mathbf{pDCF} \subset \mathbf{pCF} \subset \mathbf{pJFA} = \mathbf{JFA} \subset \mathbf{pCS}$ and $\mathbf{pREG} \subset \mathbf{pROWJ} \subset \mathbf{JFA}$. Furthermore, the family* **pROWJ** *is incomparable to* **pDCF** *and to* **pCF**. *We have $\mathbf{pROWJ} \subset \mathbf{ROWJ}$.* □

5 Closure Properties of ROWJ and pROWJ

We consider closure properties of the language families **ROWJ** and **pROWJ**. Our results are summarized in Table 1. Here we only show that **ROWJ** is *not* closed under inverse homomorphism, while the permutation closed language family **pROWJ** is closed under this operation. The proofs of the remaining closure and non-closure results will be given in the journal version of the paper.

Closed under	Language family **REG**	**pROWJ**	**ROWJ**	**JFA**
Union	yes	no	no	yes
Union with reg. lang.	yes	no	no	no
Intersection	yes	yes	no	yes
Intersection with reg. lang.	yes	no	no	no
Complementation	yes	no	no	yes
Reversal	yes	yes	no	yes
Concatenation	yes	no	no	no
Right conc. with reg. lang.	yes	no	no	no
Left conc. with reg. lang.	yes	no	no	no
Left conc. with prefix-free reg. lang.	yes	no	yes	no
Kleene star or plus	yes	no	no	no
Homomorphism	yes	no	no	no
Inv. homomorphism	yes	yes	no	yes
Substitution	yes	no	no	no
Permutation	no	yes	no	yes

Table 1. Closure properties of **ROWJ** and **pROWJ**. The gray shaded results are proven in this paper. The non-shaded closure properties for **REG** are folklore. For **ROWJ** the closure/non-closure results can be found in [3] and that for the language family **JFA** in [1, 6, 7, 12].

Theorem 14. *The family* **ROWJ** *is not closed under inverse homomorphism.*

Proof. Let A be the ROWJFA $A = (\{q_0, q_1, q_2\}, \{a, b, c\}, R, q_0, \{q_0, q_2\})$, where R consists of the rules $q_0c \to q_0$, $q_0b \to q_1$, $q_1a \to q_2$, and $q_2b \to q_1$. The ROWJFA A is depicted in Figure 2. Let $h : \{a, b\}^* \to \{a, b, c\}^*$ be the ho-

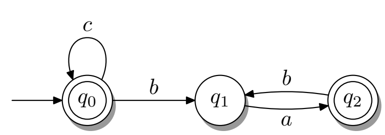

Fig. 2. The ROWJFA A satisfying $L(A) \cap \{ac, b\}^* = \{ (ac)^n b^n \mid n \geq 0 \}$.

momorphism, given by $h(a) = ac$ and $h(b) = b$. We have $h(\{a, b\}^*) = \{ac, b\}^*$.

Let now $\lambda \neq w \in L(A) \cap \{ac, b\}^*$, which implies $|w|_b > 0$. When A reads w, it reaches the first occurrence of the symbol b in state q_0. After reading this b, the automaton is in state q_1. Now, no more c can be read. So, we get $w \in (ac)^+b^+$. Whenever A is in state q_2, it has read the same number of a's and b's. This gives us $w \in \{ (ac)^n b^n \mid n > 0 \}$. That shows $L(A) \cap \{ac, b\}^* \subseteq \{ (ac)^n b^n \mid n \geq 0 \}$.

On the other hand, for $n > 0$, we have

$$q_0(ac)^n b^n \circlearrowright^n q_0 b^n a^n \circlearrowright^2 q_2 a^{n-1} b^{n-1} \circlearrowright^2 q_2 a^{n-2} b^{n-2} \circlearrowright^2 \cdots \circlearrowright^2 q_2 ab \circlearrowright^2 q_2.$$

This implies $L(A) \cap \{ac, b\}^* = \{\, (ac)^n b^n \mid n \geq 0 \,\}$. We get

$$\begin{aligned} h^{-1}(L(A)) = h^{-1}(L(A) \cap h(\{a,b\}^*)) &= h^{-1}(L(A) \cap \{ac, b\}^*) \\ &= h^{-1}(\{\, (ac)^n b^n \mid n \geq 0 \,\}) = \{\, a^n b^n \mid n \geq 0 \,\}. \end{aligned}$$

In [3] it was shown that this language is not in **ROWJ**. ◻

For the language family **pROWJ** the situation w.r.t. the closure under inverse homomorphisms is exactly the other way around.

Theorem 15. *Let Γ and Σ be alphabets and $h : \Gamma^* \to \Sigma^*$ be a homomorphism. Furthermore let $L \subseteq \Sigma^*$ be in* $\mathbf{pROWJ_n}$*, for some $n \geq 0$. Then, the language $h^{-1}(L)$ is also in* $\mathbf{pROWJ_n}$*.*

Proof. It is not difficult to see that the family of permutation closed languages is closed under inverse homomorphism. So, the language $h^{-1}(L)$ is closed under permutation. Theorem 3 gives us $|L/\sim_L| \leq n$. From $L = \bigcup_{S \in L/\sim_L} S$, we get $h^{-1}(L) = \bigcup_{S \in L/\sim_L} h^{-1}(S)$. Consider now an element $S \in L/\sim_L$, two words $v, w \in h^{-1}(S)$, and an arbitrary $u \in \Gamma^*$. Because of $h(v), h(w) \in S$, we have $h(v) \sim_L h(w)$. It follows that

$$vu \in h^{-1}(L) \Leftrightarrow h(v)h(u) \in L \Leftrightarrow h(w)h(u) \in L \Leftrightarrow wu \in h^{-1}(L).$$

We have shown $v \sim_{h^{-1}(L)} w$. So, we get $\left|h^{-1}(L)/\sim_{h^{-1}(L)}\right| \leq |L/\sim_L| \leq n$, which by Theorem 3 implies that $h^{-1}(L)$ is in $\mathbf{pROWJ_n}$. ◻

Thus we immediately get:

Corollary 16. *The family* **pROWJ** *is closed under inverse homomorphism.* ◻

6 More on Languages Accepted by ROWJFAs

In Corollary 4 a characterization of the permutation closed languages that are in **ROWJ** was given. In this section, we characterize languages in **ROWJ** for some cases where the considered language does not need to be permutation closed.

Theorem 17. *For an alphabet Σ, let $w \in \Sigma^*$ and $L \subseteq \Sigma^*$. Then, the language wL is in* **ROWJ** *if and only if L is in* **ROWJ***.*

Proof. If L is in **ROWJ**, then wL is also in **ROWJ**, because the language family **ROWJ** is closed under concatenation with prefix-free languages from the left. Now assume that wL is in **ROWJ** and $L \neq \emptyset$. We may also assume that $|w| = 1$. The general case follows from this special case *via* a trivial induction over the length of w. Thus, let $w = a$ for an $a \in \Sigma$ and let $A = (Q, \Sigma, R, s, F)$ be a DFA with $L_R(A) = aL$. In the following, we will show *via* a contradiction that the value $R(s,a)$ is defined. Assume that $R(s,a)$ is undefined and let v be an arbitrary word out of L. Because $av \in L_R(A)$, there is a symbol $b \in \Sigma_s$, two words $x \in (\Sigma \setminus \Sigma_s)^*$ and $y \in \Sigma^*$, and a state $p \in F$ such that $v = xby$ and $saxby \circlearrowright R(s,b)yax \circlearrowright^* p$. This gives us $sbyax \vdash R(s,b)yax \circlearrowright^* p$, which implies $byax \in L_R(A) = aL$. However, this is a contradiction, because $b \neq a$. So, the value $R(s,a)$ is defined.

Consider the DFA $B = (Q, \Sigma, R, R(s,a), F)$. For a word $z \in \Sigma^*$, we have $z \in L_R(B)$ if and only if $az \in L_R(A) = aL$, because of $saz \vdash R(s,a)z$. That gives us $L_R(B) = L$ and we have shown that L is in **ROWJ**. $\square$

From the previous theorem and Corollary 4 we get:

Corollary 18. *For an alphabet Σ, let $w \in \Sigma^*$ and let $L \subseteq \Sigma^*$ be a permutation closed language. Then, the set wL is in* **ROWJ** *if and only if the Myhill-Nerode relation $\sim_L$ has only a finite number of positive equivalence classes.* $\square$

Next, we will give a characterization for the concatenation Lw of a language L and a word w. To do so, we need the following lemma. It treats the case of an ROWJFA that is only allowed to jump over one of the input symbols.

Lemma 19. *Let $A = (Q, \Sigma, R, s, F)$ be a DFA with a symbol $a \in \Sigma$ such that $R(q,b)$ is defined for all $(q,b) \in Q \times (\Sigma \setminus \{a\})$. Then, $L_R(A)$ is regular.* $\square$

Our characterization for languages of the form Lw generalizes a result from [3], which says that the language $\{\, va \mid v \in \{a,b\}^*, |v|_a = |v|_b \,\}$ is not in **ROWJ**:

Theorem 20. *For an alphabet Σ, let $w \in \Sigma^*$ be a non-empty word and $L \subseteq \Sigma^*$. Then, the language Lw is in* **ROWJ** *if and only if L is regular.* $\square$

Now, we consider the case of two languages over disjoint alphabets.

Theorem 21. *For disjoint alphabets Σ_1 and Σ_2, let $L_1 \subseteq \Sigma_1^*$ and $L_2 \subseteq \Sigma_2^*$ with $L_1 \neq \emptyset \neq L_2 \neq \{\lambda\}$ such that L_1L_2 is in* **ROWJ**. *Then, the language L_1 is regular and L_2 is in* **ROWJ**. $\square$

Adding prefix-freeness for L_1, we get an equivalence, by Theorem 21 and the closure of **ROWJ** under left-concatenation with prefix-free regular sets.

Corollary 22. *For disjoint alphabets Σ_1 and Σ_2, let $L_1 \subseteq \Sigma_1^*$ be a prefix-free language and $L_2 \subseteq \Sigma_2^*$ be an arbitrary language with $L_1 \neq \emptyset \neq L_2 \neq \{\lambda\}$. Then, the language L_1L_2 is in* **ROWJ** *if and only if L_1 is regular and L_2 is in* **ROWJ**. $\square$

The previous corollary directly implies the following characterization:

Corollary 23. *For disjoint alphabets Σ_1 and Σ_2, let $L_1 \subseteq \Sigma_1^*$ be a prefix-free language and $L_2 \subseteq \Sigma_2^*$ be a permutation closed language with $L_1 \neq \emptyset \neq L_2 \neq \{\lambda\}$. Then, the language L_1L_2 is in* **ROWJ** *if and only if L_1 is regular and the Myhill-Nerode relation $\sim_{L_2}$ has only a finite number of positive equivalence classes.* $\square$

If a non-empty language and a non-empty permutation closed language over disjoint alphabets are separated by a symbol, we get the following result:

Corollary 24. *For disjoint alphabets Σ_1 and Σ_2, let $L_1 \subseteq \Sigma_1^*$ be a non-empty language and $L_2 \subseteq \Sigma_2^*$ be a non-empty permutation closed language. Furthermore, let $a \in \Sigma_2$. Then, the language $L_1 a L_2$ is in* **ROWJ** *if and only if L_1 is regular and the Myhill-Nerode relation $\sim_{L_2}$ has only a finite number of positive equivalence classes.* □

For an alphabet $\Sigma = \{a_1, a_2, \ldots, a_n\}$, the family of subsets of $a_1^* a_2^* \ldots a_n^*$ is kind of a counterpart of the family of permutation closed languages over Σ. In a language $L \subseteq a_1^* a_2^* \ldots a_n^*$, for each word $w \in L$, no other permutation of w is in L. We can characterize the subsets of $a_1^* a_2^* \ldots a_n^*$ that are in **ROWJ**.

Theorem 25. *Let $\{a_1, a_2, \ldots, a_n\}$ be an alphabet and $L \subseteq a_1^* a_2^* \ldots a_n^*$. Then, the language L is in* **ROWJ** *if and only if L is regular.* □

References

1. S. Beier, M. Holzer, and M. Kutrib. Operational state complexity and decidability of jumping finite automata. In Proc. 21st DLT, number 10396 in LNCS, pages 96–108, Liège, Belgium, 2017. Springer.
2. S. Bensch, H. Bordihn, M. Holzer, and M. Kutrib. On input-revolving deterministic and nondeterministic finite automata. *Inform. Comput.*, 207(11):1140–1155, 2009.
3. H. Chigahara, S. Fazekas, and A. Yamamura. One-way jumping finite automata. *Internat. J. Found. Comput. Sci.*, 27(3):391–405, 2016.
4. C. Culy. Formal properties of natural language and linguistic theories. *Linguist. Philos.*, 19:599–617, 1996.
5. L. E. Dickson. Finiteness of the odd perfect and primitive abundant numbers with n distinct prime factors. *American J. Math.*, 35(4):413–422, 1913.
6. H. Fernau, M. Paramasivan, and M. L. Schmid. Jumping finite automata: Characterizations and compliexity. In Proc. 20th CIAA, number 9223 in LNCS, pages 89–101, Umeå, Sweden, 2015. Springer.
7. H. Fernau, M. Paramasivan, M. L. Schmid, and V. Vorel. Characterization and complexity results on jumping finite automata. http://arxiv.org/abs/1512.00482, 2015.
8. Ginsburg, S., Spanier, E.H.: Bounded ALGOL-like languages. Trans. AMS 113: 333–368, 1964
9. S. A. Greibach. An infinite hierarchy of context-free languages. *J. ACM*, 16(1):91–106, 1969.
10. J. E. Hopcroft and J. D. Ullman. *Introduction to Automata Theory, Languages and Computation.* Addison-Wesley, 1979.
11. A. Meduna and P. Zemek. Jumping finite automata. *Internat. J. Found. Comput. Sci.*, 23(7):1555–1578, 2012.
12. A. Meduna and P. Zemek. *Regulated Grammars and Automata*, chapter 17: Jumping Finite Automata, pages 567–585. Springer, 2014.
13. V. Vorel. Basic properties of jumping finite automata. http://arxiv.org/abs/1511.08396v2, 2015.