Factorizations and Universal Automaton of Omega Languages

Vincent Carnino, Sylvain Lombardy

HAL Id: hal-00841873

https://hal-upec-upem.archives-ouvertes.fr/hal-00841873

Submitted on 9 Jul 2013

Factorizations and Universal Automaton of omega Languages

Vincent Carnino[1] and Sylvain Lombardy[2]

[1] LIGM, Université Paris-Est Marne-la-Vallée
[2] LABRI, Université de Bordeaux - IPB

Abstract. In this paper, we extend the concept of factorization on finite words to ω-rational languages and show how to compute them. We define a normal form for Büchi automata and introduce a universal automaton for Büchi automata in normal form. We prove that, for every ω-rational language, this Büchi automaton, based on factorization, is canonical and that it is the smallest automaton that contains the morphic image of every equivalent Büchi automaton in normal form.

1 Introduction

When considering rational languages on finite words, different kinds of formalism may be studied: automata, semigroups, rational expressions, *etc.* There exist similar notions for rational languages on infinite words, also called ω-rational languages, which are a rational extension of languages on finite words. Indeed, classical semigroups and rational expressions have been extended to ω-semigroups and ω-rational expressions respectively. For the automata counterpart, there is not a unique approach but several ones depending on the acceptance mode: Büchi automata, Muller automata, Rabin automata, Streett automata, *etc.* In this paper, we focus on Büchi automata, which are the most intuitive kind of automata accepting infinite words.

Infinite words are widely used in computer science, especially in order to describe the behaviour of infinite processes. Some well-known results on finite words are also transposable to infinite words, like Kleene's theorem which states that ω-rational languages are exactly languages that are recognized by finite Büchi automata and, as an extention, by finite ω-semigroups. Yet there are other properties that are not transposable to infinite words like the existence of a canonical deterministic minimal automaton: deterministic Büchi automata are strictly less expressive than nondeterministic ones and, as a consequence, there is no notion of minimal Büchi automaton. Carton and Michel [1] have proved that *prophetic automata* (which are the pertinent notion for "right-left" determinism) are as expressive as Büchi automata, but there is no unique minimal prophetic automata for some ω-rational languages.

Yet, the minimal automaton is not the only canonical automaton associated to a language. In 1971, Conway [2] introduced the notions of factorization and factor matrix of a rational language. This concept has led to the definition of

a new object called the universal automaton of a language [5, 9]. It has many significant properties since any automaton that recognizes a language L has a morphic image which is a subautomaton of the universal automaton of L. For example, it may be used to compute a NFA of minimum size [8], or in theoretical proofs for the existence of automata with specific properties (star height [4], reversibility [3], *etc.*).

We extend the concept of factorization on finite words to ω-rational languages. Using these ω-factorizations, we build the universal automaton of an ω-rational language. We prove that, up to a conversion to a normal form, every Büchi automaton has a morphic image in this automaton (universal property), and that this automaton is minimal for this property.

In the first part, we give some basic definitions about languages (on finite and infinite words) and about automata that will be used in the course of this article. We recall some basic notions such as the past and the future of states in both NFA and Büchi automata and some new ones like the normal form of a Büchi automaton. Finally, we recall the definition of ω-semigroups given by [6] and the principle of ω-rational language recognition by ω-semigroup.

In the second part, we define ω-factorizations and pure ω-factorizations which will both be used to define the universal automaton. Then we explain how to compute them using the transition ω-semigroup of a Büchi automaton. For ω-factorizations, we describe another computation based on prophetic automata.

The last part is devoted to the definition of the universal automaton of an ω-rational language $\mathcal{L}$. It involves both pure and standard ω-factorizations as well as positive factorizations on finite words. Finally, we state the main properties of this automaton: it accepts exactly $\mathcal{L}$, has the universal property and is minimal among universal automata for $\mathcal{L}$.

2 Definitions

2.1 Languages and ω-Languages

Classically A^* denotes the free monoid generated by the alphabet A. The length of a word w in A^* is $|w|$, and for $i \in [1; |w|]$, w_i is the i-th letter of w. We denote the empty word by ε and A^+ denotes the free semigroup of non empty words.
The product over A^* naturally extends to languages. The Kleene star L^* of a language L is defined as the union of powers of L, while L^+ is the union of positive powers of L.
The set of infinite words (or ω-words) over A is A^ω; the mixed product of a word u and an ω-word v is the unique ω-word w which can be factorized into a prefix u and a suffix v. A subset of A^ω is called an ω-language. Like the concatenation product, the mixed product naturally extends to a product between languages and ω- languages.
For every word u in A^+, the ω-word u^ω is the infinite repetition of u. If L is a subset of A^+, the ω-language L^ω is the set of infinite concatenations of words of L.

These operations lead to the classical definitions of rational and ω-rational languages.

Definition 1. *The set of rational languages over A is the smallest set of languages which contains every finite language over A and which is closed under union, concatenation and iteration.*

Definition 2. *A set $\mathcal{L}$ of ω-words is an ω-power language if there exists a rational language $K \subseteq A^+$ such that $\mathcal{L} = K^\omega$.*
The set of ω-rational languages over A is the smallest set which contains the empty set, every ω-power language over A, and which is closed under mixed product with rational languages and union.

It is straightforward, that for every ω-rational language $\mathcal{L}$, there exists a finite set of pairs of rational languages $(X_i, Y_i)_{i \in I}$ such that

$$\mathcal{L} = \bigcup_{i \in I} X_i Y_i^\omega. \tag{1}$$

2.2 Automata

An automaton is a 5-tuple $\mathcal{A} = (Q, A, E, I, F)$, where Q is a finite set of states, A is an alphabet, $E \subseteq Q \times A \times Q$ is the set of transitions, $I \subseteq Q$ is the set of initial states and $F \subseteq Q$ is the set of final states. If (p, a, q) is an element of the set of transitions, we denote it $p \xrightarrow{a} q$. As usual, such an automaton can be considered as a labeled graph and we use the graph terminology. A path is a finite sequence of consecutive transitions. An ω-path is an infinite sequence of consecutive transitions. The label of a path is the sequence of the labels of its transitions.

A morphism of automata from $\mathcal{A} = (Q, A, E, I, F)$ into $\mathcal{B} = (R, A, G, J, T)$ is a mapping φ from Q to R such that $\varphi(I) \subseteq J$, $\varphi(F) \subseteq T$, and $\varphi(p) \xrightarrow{a} \varphi(q)$ is in G for every $p \xrightarrow{a} q$ in E.

Definition 3. *Let $\mathcal{A}$ be an automaton. Let p be a state of $\mathcal{A}$. The past of p, $\mathsf{Past}_\mathcal{A}(p)$, is the language of words that label a path from some initial state to p. The set of words that label a path between two states p and q is denoted by $\mathsf{Trans}_\mathcal{A}(p, q)$.*

The semantic of an automaton depends on the acceptance mode. In this paper we only consider two kinds of automata: NFA and Büchi automata; their acceptance mode is described in the two following definitions.

Definition 4. *Let $\mathcal{A}$ be a NFA. A path of $\mathcal{A}$ is accepting if it ends at a final state. The future of a state p, $\mathsf{Fut}_\mathcal{A}(p)$, is the set of words that label accepting paths starting at state p. The (rational) language recognized by $\mathcal{A}$ is the union of futures of the initial states of $\mathcal{A}$.*

Definition 5. *Let $\mathcal{A}$ be a Büchi automaton. An ω-path is accepting if it meets an infinite number of occurrences of final states. The future of a state p, $\mathsf{Fut}_{\mathcal{A}}(p)$, is the set of ω-words that label accepting ω-paths starting at state p. The ω-rational language recognized by $\mathcal{A}$ is the union of futures of the initial states of $\mathcal{A}$.*

We introduce here a normal form for Büchi automata. Indeed, in ω-rational languages, the acceptation of a word is the conjunction of conditions on the finite prefixes of the word and conditions dealing with the infinite behaviour. The normal form we consider here consists in splitting the automaton in two parts: one *transient* part in which finite prefixes are read, and *final* components which process the infinite part.

Definition 6. *A state of Büchi automaton $\mathcal{A}$ is* transient *if it is not accessible from a final state. A strongly connected component (SCC) is* final *if it contains a* unique *final state, at least one transition, and if every state accessible from this final state is in the same SCC. A Büchi automaton is in* normal form *if*
i) every initial state is transient;
ii) every final state is in a final SCC;
iii) for every non final state q of a final SCC S, every predecessor of q is in S.

Definition 7. *Let $\mathcal{A} = (Q, A, E, I, F)$ be a Büchi automaton. Let G be the set of final states of $\mathcal{A}$ that belong to a non trivial SCC, and let $\mathcal{S}$ be the function which maps every state f of G onto the SCC of f. The normalization of $\mathcal{A}$ is the automaton $\mathcal{A}_{\mathsf{nf}} = (Q', A, E', I', F')$ defined by :*

- $Q' = Q \cup \{(p, f) \mid f \in G, p \in \mathcal{S}(f)\}$;
- $I' = I$, $F' = \{(f, f) \mid f \in G\}$;
- $E' = E \cup \{p \xrightarrow{a} (f, f) \mid p \xrightarrow{a} f \in E, f \in G\}$
  $\cup \{(p, f) \xrightarrow{a} (q, f) \mid p \xrightarrow{a} q \in E, f \in G, p, q \in \mathcal{S}(f)\}$.

Proposition 1. *Let $\mathcal{A}$ be a Büchi automaton. The normalization $\mathcal{A}_{\mathsf{nf}}$ of $\mathcal{A}$ is an equivalent Büchi automaton in normal form.*

Example 1. Figure 1 shows a Büchi automaton and its normal form.

2.3 Semigroup Recognition

Rational languages can be defined by finite automata and rational expressions, but they can also be characterized by morphisms into finite monoids. A semigroup is a set endowed with an associative product, a monoid is a semigroup with a unit element. If S is a semigroup, S^1 is the monoid obtained by adding a unit element. A morphism of semigroups is a mapping between two semigroups which commutes with the product. A morphism of monoids also preserves the unit element. A language K of A^* is said to be *recognizable* if there exists a finite monoid M and a surjective morphism φ from A^* onto M such that $K = \varphi^{-1}(\varphi(K))$; we then say that K is recognized by (M, φ). The recognizable languages of A^* are exactly the rational languages.
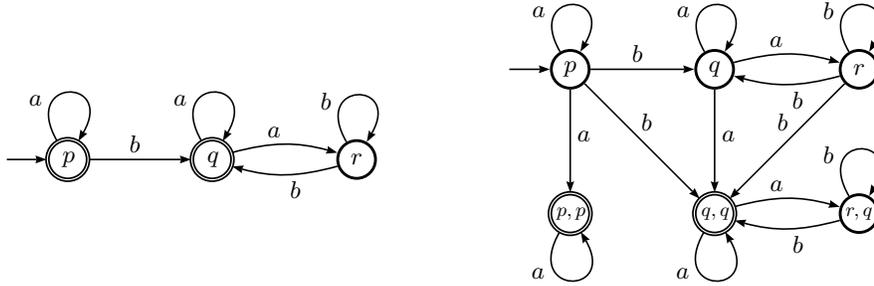
Fig. 1. A Büchi automaton and its normal form.

For languages in A^+, there exists a similar notion of recognizability by finite semigroup. Clearly, a language L of A^* is recognized by a finite monoid if and only if $L \cap A^+$ is recognized by a finite semigroup.

This approach has turned out to be very fruitful and there have been many attempts to extend it to ω-rational languages. It appears that the appropriate notion of recognizability of ω-rational languages by a finite algebraic structure requires to embed them in a larger structure, called ω-*semigroup*. The recognizability theorem [6, 7] (Theorem 1) requires a few definitions.

Definition 8. *An ω-semigroup is a pair $S = (S_+, S_\omega)$, where S_+ is a semigroup, endowed with:*

- *A* mixed product $S_+ \times S_\omega \to S_\omega$ *such that for every (u, v, w) in $S_+ \times S_+ \times S_\omega$, $(uv)w = u(vw)$.*
- *A surjective* infinite product $S_+^\omega \to S_\omega$ *which is compatible with finite associativity and mixed product.*

Definition 9. *Let $S = (S_+, S_\omega)$ and $T = (T_+, T_\omega)$ be two ω-semigroups. A* morphism of ω-semigroups *is a pair $\varphi = (\varphi_+, \varphi_\omega)$ consisting of a semigroup morphism $\varphi_+ : S_+ \to T_+$ and a mapping $\varphi_\omega : S_\omega \to T_\omega$ that commutes with the infinite product.*

These notions allow to define the recognizability by ω-semigroup.

Definition 10. *Let T be an ω-semigroups and let $\varphi : A^\infty = (A^+, A^\omega) \to T$ be a surjective morphism of ω-semigroups. A set $\mathcal{L}$ of ω-words over A is* recognized by (T, φ) *if $\mathcal{L} = \varphi_\omega^{-1}(\varphi_\omega(\mathcal{L}))$.*
A set of ω-words over A is recognizable *if there exists a finite ω-semigroup T and a surjective morphism φ such that $\mathcal{L}$ is* recognized by (T, φ).

Theorem 1. *[6] A set $\mathcal{L}$ of ω-words over A is an ω-rational language if and only if it is* recognizable.

3 Factorizations of Languages

3.1 Definitions and Properties

The factorizations of languages were introduced by Conway in [2]. He has shown that every rational language has a finite number of maximal factorizations and that this property characterizes rational languages.

Definition 11. *Let L be a language over A. A factorization of L is a pair of languages $X = (X_1, X_2)$ such that $X_1 X_2$ is a subset of L. A positive factorization $X = (X_1, X_2)$ is a factorization such that X_1 and X_2 are non empty subsets of A^+. A factorization (resp. positive factorization) X is maximal if for every factorization (resp. positive factorization) X' of L, $X_1 \subseteq X_1'$ and $X_2 \subseteq X_2'$ implies $X = X'$. We denote by $\mathfrak{F}(L)$ the set of maximal factorizations of L and by $\mathfrak{F}_+(L)$ the set of maximal positive factorizations of L.*

We extend this definition to ω-semigroups; in this case, we define two kinds of factorizations: the ω-factorizations which are the straightforward extension of factorizations on finite words and the pure ω-factorizations that involve the infinite iteration.

Definition 12. *Let $S = (S_+, S_\omega)$ be an ω-semigroup and let K be a subset of S_ω. An ω-factorization of K is a pair $X = (X_1, X_2)$, with $X_1 \subseteq S_+^1$ and $X_2 \subseteq S_\omega$ such that $X_1 X_2$ is a subset of K. A pure ω-factorization of K is a pair $Y = (Y_1, Y_2)$, with $Y_1, Y_2 \subseteq S_+$ and $Y_2 \neq \emptyset$, such that $Y_1 Y_2^\omega$ is a subset of K. An ω-factorization (resp. pure ω-factorization) X of K is maximal if for every ω-factorization (resp. pure ω-factorization) Y of K, then $X_1 \subseteq Y_1$ and $X_2 \subseteq Y_2$ implies $X = Y$. We define $\mathfrak{F}(K)$ as the set of maximal ω-factorizations of K and $\mathfrak{F}_p(K)$ as the set of maximal pure ω-factorizations of K.*

If X is a maximal ω-factorization of K then $X_1 = \{x \in S_+^1 \mid xX_2 \subseteq K\}$, and $X_2 = \{y \in S_\omega \mid X_1 y \subseteq K\}$. If Y is a maximal pure ω-factorization of K then $Y_1 = \{x \in S_+ \mid xY_2^\omega \subseteq K\}$ and $Y_2^+ = Y_2$, but Y_2 is not characterized by Y_1.

Example 2. We consider the language $L_1 = A^*(aa + bb)A^*$ and the ω-regular language $\mathcal{L}_1 = L_1^\omega$. The pairs $(A^*, \mathcal{L}_1)$ and $(\emptyset, A^\omega)$ are the two maximal ω-factorizations of $\mathcal{L}_1$. The pairs $(A^+, L_1 + aA^*a + a)$ and $(A^+, L_1 + bA^*b + b)$ are maximal pure ω-factorizations of $\mathcal{L}_1$.[3]

Proposition 2. *Let $\mathcal{L}$ be an ω-rational language. If Y is in $\mathfrak{F}_p(\mathcal{L})$, then there exists X in $\mathfrak{F}(\mathcal{L})$ such that $Y_1 = X_1 \cap A^+$ and $Y_2^\omega \subseteq X_2$.*

Proposition 3. *Let $\mathcal{L}$ be an ω-rational language over A. Let T be a finite ω-semigroup and $\varphi : A^\infty \to T$ such that $\mathcal{L}$ is recognized by (T, φ). Then, for every X in $\mathfrak{F}(\mathcal{L})$, the set $X_1 \cap A^+$ is recognized by (T_+, φ_+), and X_2 is recognized by (T, φ), and for every Y in $\mathfrak{F}_p(\mathcal{L})$, both Y_1 and Y_2 are recognized by (T_+, φ_+).*

---

[3] Notice that the image of a maximal pure ω-factorizations in the syntactic ω-semigroup is not necessarily a union of *linked pairs* (*cf.* [6]); in $(A^+, L_1 + aA^*a + a)$, the image of a is not an idempotent.

Proof. Let $\mathcal{L}$ be an ω-rational language and let $\varphi : A^\infty \to T$ be a morphism of ω-semigroups that recognizes $\mathcal{L}$, where T is a finite ω-semigroup. Let φ_* be the morphism of monoids from A^* onto T_+^1 which is the natural extention of φ_+.

Let $X \in \mathfrak{F}(\mathcal{L})$, then $\varphi_*(X_1)\varphi_\omega(X_2) \subseteq \varphi_\omega(\mathcal{L})$. Therefore, there exists α in $\mathfrak{F}(\varphi(\mathcal{L}))$ such that $\varphi_*(X_1) \subseteq \alpha_1$ and $\varphi_\omega(X_2) \subseteq \alpha_2$. Since $\mathcal{L} = \varphi_\omega^{-1}(\varphi_\omega(\mathcal{L}))$, then $\varphi_*^{-1}(\alpha_1)\varphi_\omega^{-1}(\alpha_2) \subseteq \mathcal{L}$, thus by maximality of X, it holds $X_1 = \varphi_*^{-1}(\alpha_1)$ and $X_2 = \varphi_\omega^{-1}(\alpha_2)$: the maximal factorizations of $\mathcal{L}$ are recognized by φ.

Let Y be in $\mathfrak{F}_p(\mathcal{L})$. The pair $\gamma = (\varphi_+(Y_1), \varphi_+(Y_2))$ is a pure ω-factorization of $\varphi_\omega(\mathcal{L})$, thus there exists β in $\mathfrak{F}_p(\varphi_\omega(\mathcal{L}))$ such that $\gamma_1 \subseteq \beta_1$ and $\gamma_2 \subseteq \beta_2$. Therefore $Y_1 \subseteq \varphi_+^{-1}(\beta_1)$ and $Y_2 \subseteq \varphi_+^{-1}(\beta_2)$, and, since $\varphi_+^{-1}(\beta_1)(\varphi_+^{-1}(\beta_2))^\omega \subseteq \varphi_\omega^{-1}(\varphi_\omega(\mathcal{L})) = \mathcal{L}$, by maximality of Y, it holds $Y_1 = \varphi_+^{-1}(\beta_1)$ and $Y_2 = \varphi_+^{-1}(\beta_2)$. The maximal pure ω-factorizations of $\mathcal{L}$ are therefore recognized by φ. $\square$

Corollary 1. *Let $\mathcal{L}$ be an ω-rational language over A. The sets $\mathfrak{F}(\mathcal{L})$ and $\mathfrak{F}_p(\mathcal{L})$ are finite and each of their elements is a pair of (ω-)rational languages. Moreover,*

$$\mathcal{L} = \bigcup_{X \in \mathfrak{F}(\mathcal{L})} X_1 X_2 = \bigcup_{Y \in \mathfrak{F}_p(\mathcal{L})} Y_1 Y_2^\omega \tag{2}$$

Corollary 2. *Let $\mathcal{A}$ be a Büchi automaton and let $\mathcal{L}$ be the ω-rational language accepted by $\mathcal{A}$. The sets $\mathfrak{F}(\mathcal{L})$ and $\mathfrak{F}_p(\mathcal{L})$ are effectively computable.*

Proof. The transition ω-semigroup $S = (S_+, S_\omega)$ of $\mathcal{A}$ is computable (*cf.* [6]), as well as the morphism φ such that $\mathcal{L}$ is recognized by (S, φ). By Proposition 3, every computation of maximal factorizations can be done in S, which is finite. $\square$

This proof induces a bound on the number of maximal ω-factorizations. Since each maximal ω-factorization is characterized by one of its factors, the number of ω-factorizations is at most $\min(2^{|S_+|+1}, 2^{|S_\omega|})$. Likewise, the number of maximal pure ω-factorizations is at most $2^{|S_+|}$.

In contrast with the case of finite words, the finiteness of $\mathfrak{F}(\mathcal{L})$ and $\mathfrak{F}_p(\mathcal{L})$ does not imply that $\mathcal{L}$ is ω-rational.

Example 3. Let $\mathcal{L}_2 = \{\prod_{i \geqslant 0} a^{f(i)}b \mid f : \mathbb{N} \to \mathbb{N}$ and $\forall i, f^{-1}(i)$ is finite$\}$; it is not ω-rational and, yet, $\mathfrak{F}(\mathcal{L}_2) = \{(A^*, \mathcal{L}_2), (\emptyset, A^\omega)\}$ and $\mathfrak{F}_p(\mathcal{L}_2) = \{(\emptyset, A^+)\}$.

3.2 Computation of Maximal ω-Factorizations from Prophetic Automata

In this part, we present an alternative computation of the factorizations of an ω-rational language based on a *prophetic* automaton recognizing this language.

Definition 13. *[1] A Büchi automaton $\mathcal{A}$ over A is prophetic if the future of every state is non empty and the futures of states are pairwise disjoint.*

Theorem 2. *[1] Every ω-rational language can be recognized by a prophetic automaton.*

The conversion of a Büchi automaton into a prophetic automaton is effective, but a bit complicated. Nevertheless, some natural ω-rational languages have very simple prophetic automata.

To compute the maximal factorizations from a prophetic automaton, we use the well-known *subset construction*. The aim is not to obtain a deterministic equivalent Büchi automata (which may not exist), but to compute a set of states, that will be used in Proposition 4 to characterize maximal ω-factorizations.

If $\mathcal{A} = (Q, A, E, I, F)$ is a (Büchi) automaton, for each word u in A^*, the set of accessible states by u from a subset K of Q is $\delta_{\mathcal{A}}(K, u) = \{q \mid \exists p \in K \text{ and } u \in \mathsf{Trans}_{\mathcal{A}}(p, q)\}$.

The *subset construction* of $\mathcal{A}$ is the set $P = \{K \subseteq Q \mid \exists u \in A^*, \; K = \delta_{\mathcal{A}}(I, u)\}$. Notice that $I = \delta_{\mathcal{A}}(I, \varepsilon)$ is always in P and that P can be incrementaly computed, since $\delta_{\mathcal{A}}(I, ua) = \delta_{\mathcal{A}}(\delta_{\mathcal{A}}(I, u), a)$.

Proposition 4. *Let $\mathcal{A}$ be a prophetic automaton with set of states Q recognizing $\mathcal{L} \subseteq A^\omega$. Let P be the subset construction of $\mathcal{A}$ and let $P_\cap$ be the smallest set containing the element Q, every element of P, and closed under intersection. We set, for every K in $P_\cap$, and every X in $\mathfrak{F}(\mathcal{L})$,*

$$\varphi(K) = \left(\bigcap_{p \in K} \mathsf{Past}_{\mathcal{A}}(p), \; \bigcup_{p \in K} \mathsf{Fut}_{\mathcal{A}}(p) \right), \qquad \psi(X) = \bigcap_{u \in X_1} \delta_{\mathcal{A}}(I, u). \qquad (3)$$

Then, φ is a one-to-one mapping between $P_\cap$ and $\mathfrak{F}(\mathcal{L})$ and $\psi = \varphi^{-1}$.

Proof. Let $X \in \mathfrak{F}(\mathcal{L})$ and let $H = \psi(X)$, which is clearly in $P_\cap$, since for each word u, the set $\delta_{\mathcal{A}}(I, u)$ is in P (if $X_1 = \emptyset$, then $H = Q$).

We prove now that the factorization $Y = \varphi(H)$ is equal to X. It holds $H = \bigcap_{u \in X_1} \delta_{\mathcal{A}}(I, u) = \{p \in Q \mid X_1 \subseteq \mathsf{Past}_{\mathcal{A}}(p)\}$, thus $X_1 \subseteq \bigcap_{p \in H} \mathsf{Past}_{\mathcal{A}}(p) = Y_1$. For each v in X_2, there exists a unique p in Q such that v is in $\mathsf{Fut}_{\mathcal{A}}(p)$; it holds $X_1 v \subseteq \mathcal{L}$, hence $X_1 \subseteq \mathsf{Past}_{\mathcal{A}}(p)$ and p is in H. Thus, $X_2 \subseteq \bigcup_{p \in H} \mathsf{Fut}_{\mathcal{A}}(p)$, and $X = Y$ by maximality of X.

Conversely, let K be in $P_\cap$, and $Y = \varphi(K)$. It holds

$$\begin{aligned} K &= \bigcap \{\delta_{\mathcal{A}}(I, u) \mid u \in A^*, K \subseteq \delta_{\mathcal{A}}(I, u)\} \\ &= \bigcap \{\delta_{\mathcal{A}}(I, u) \mid u \in \bigcap_{p \in K} \mathsf{Past}_{\mathcal{A}}(p)\} = \bigcap_{u \in Y_1} \delta(I, u). \end{aligned} \qquad (4)$$

Let X be in $\mathfrak{F}(\mathcal{L})$ such that $Y_1 \subseteq X_1$ and $Y_2 \subseteq X_2$. Since $Y_1 \subseteq X_1$, it holds $\psi(X) \subseteq K$. For each p in K, there exists v in $\mathsf{Fut}_{\mathcal{A}}(p) \subseteq Y_2 \subseteq X_2$. Thus $X_1 v \subseteq \mathcal{L}$ and p is in $\bigcap_{u \in X_1} \delta_{\mathcal{A}}(I, u) = \psi(X)$. Therefore, $\psi(X) = K$ and $Y = \varphi(\psi(X)) = X$. $\qquad \square$

4 Universal Automaton

In this part, we extend the definition of the universal automaton [5] of a language to infinite words. In the case of finite words, the universal automaton of L is the

smallest automaton which recognizes L and in which every equivalent automaton has a morphic image.

We shall first describe the definition of the universal automaton of $\mathcal{L}$ and then prove that it is actually the smallest Büchi automaton which recognizes $\mathcal{L}$ and in which every equivalent automaton in normal form has a morphic image.

4.1 Definition of the Universal Automaton

The definition of the universal automaton of an ω-rational language involves ω-factorizations, pure ω-factorizations and positive factorizations.

Definition 14. *Let $\mathcal{L}$ be an ω-rational language on A. For each Y in $\mathfrak{F}_p(\mathcal{L})$, we set $\mathcal{Z}_Y = \{Z \in \mathfrak{F}_+(Y_2) \mid Z \neq (Y_2, Y_2)\}$. The universal automaton $\mathcal{U}_\mathcal{L}$ of $\mathcal{L}$ is a Büchi automaton defined as follows.*
The set of states of $\mathcal{U}_\mathcal{L}$ is the union of $\mathfrak{F}(\mathcal{L})$, $\mathfrak{F}_p(\mathcal{L})$ and $\mathcal{Z}_Y$ for each $Y \in \mathfrak{F}_p(\mathcal{L})$.
The set of final states of $\mathcal{U}_\mathcal{L}$ is $\mathfrak{F}_p(\mathcal{L})$.
The set of initial states of $\mathcal{U}_\mathcal{L}$ is $\{X \in \mathfrak{F}(\mathcal{L}) \mid \varepsilon \in X_1\}$.
The set of transitions of $\mathcal{U}_\mathcal{L}$ is

$$\{X \xrightarrow{a} X' \mid a \in A, \ X, X' \in \mathfrak{F}(\mathcal{L}) \ and \ X_1 a \subseteq X'_1\}$$
$$\cup \{X \xrightarrow{a} Y \mid a \in A, X \in \mathfrak{F}(\mathcal{L}), Y \in \mathfrak{F}_p(\mathcal{L}), \ and \ X_1 a Y_2^\omega \subseteq \mathcal{L}\}$$
$$\cup \{K \xrightarrow{a} Y \mid a \in A, \ Y \in \mathfrak{F}_p(\mathcal{L}), K \in \mathcal{Z}_Y \cup \{Y\}, a \in K_2\}$$
$$\cup \{K \xrightarrow{a} Z \mid a \in A, \ \exists Y \in \mathfrak{F}_p(\mathcal{L}), K \in \mathcal{Z}_Y \cup \{Y\}, Z \in \mathcal{Z}_Y \ and \ a Z_2 \subseteq K_2\}.$$

Example 4. Let $\mathcal{L}_b$ be the language of words with an infinite number of 'b', which is recognized by the automaton of Figure 2 (left). The maximal ω-factorizations and maximal pure ω-factorizations are $\mathfrak{F}(\mathcal{L}_b) = \{(A^*, \mathcal{L}_b), (\emptyset, A^\omega)\}$ and $\mathfrak{F}_p(\mathcal{L}_b) = \{(\emptyset, A^+), (A^+, K_b)\}$, where $K_b = A^* b A^*$. From the right factors of these maximal pure ω-factorizations, we obtain the following sets of maximal positive factorizations: $\mathfrak{F}_+(A^+) = \{(A^+, A^+)\}$ and $\mathfrak{F}_+(K_b) = \{(A^+, K_b), (K_b, A^+)\}$. Notice that $\mathcal{Z}_{(\emptyset, A^+)} = \mathfrak{F}_+(A^+) \setminus \{(A^+, A^+)\} = \emptyset$.
The construction of the universal automaton of $\mathcal{L}_b$ follows; it is shown in Figure 2 (right).

4.2 Basic Properties of the Universal Automaton

The choices concerning the definition of $\mathcal{U}_\mathcal{L}$ could seem arbitrary at first glance. Nevertheless, the following propositions, which seem to naturally follow from the definition of the universal automaton, and that are required for the soundness of this notion, may not be true as soon as the definition of the universal automaton is slightly modified.

The conditions that define the transitions of the universal automaton can be generalized to charaterize the paths in this automaton.
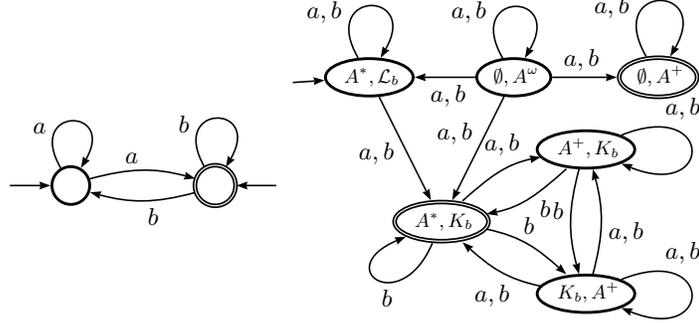
Fig. 2. A prophetic Büchi automaton and the universal automaton recognizing words with an infinite number of 'b'.

Lemma 1. *Let $\mathcal{L}$ be an ω-rational language. Let $\mathcal{U}_\mathcal{L}$ be the universal automaton of $\mathcal{L}$ and let $\mathcal{Z}_Y$ be defined as in Definition 14. Let w be a non empty (finite) word. Let X and X' in $\mathfrak{F}(\mathcal{L})$, Y in $\mathfrak{F}_p(\mathcal{L})$, Z in $\mathcal{Z}_Y$, and K in $\mathcal{Z}_Y \cup Y$. It holds:*

a) $w \in \mathsf{Trans}_{\mathcal{U}_\mathcal{L}}(X, X') \Leftrightarrow X_1 w \subseteq X'_1$;   *b)* $w \in \mathsf{Trans}_{\mathcal{U}_\mathcal{L}}(X, Y) \Leftrightarrow X_1 w Y_2{}^\omega \subseteq \mathcal{L}$;

   c) $w \in \mathsf{Trans}_{\mathcal{U}_\mathcal{L}}(K, Y) \Leftrightarrow w \in K_2$;   *d)* $w \in \mathsf{Trans}_{\mathcal{U}_\mathcal{L}}(K, Z) \Leftrightarrow w Z_2 \subseteq K_2$.

The future and the past of states of the universal automaton are closely related to the factorizations which define them.

Proposition 5. *Let $\mathcal{L}$ be an ω-rational language and let $\mathcal{U}_\mathcal{L}$ be the universal automaton of $\mathcal{L}$.*
1. For every X in $\mathfrak{F}(\mathcal{L})$,

$$a)\ \mathsf{Past}_{\mathcal{U}_\mathcal{L}}(X) = X_1 \qquad b)\ \mathsf{Fut}_{\mathcal{U}_\mathcal{L}}(X) = X_2.$$

2. For every Y in $\mathfrak{F}_p(\mathcal{L})$ and for every $Z \in \mathcal{Z}_Y$,

$$a)\ \mathsf{Past}_{\mathcal{U}_\mathcal{L}}(Y) = Y_1 \qquad b)\ \mathsf{Trans}_{\mathcal{U}_\mathcal{L}}(Y, Y) = Y_2 \quad c)\ \mathsf{Fut}_{\mathcal{U}_\mathcal{L}}(Y) = Y_2{}^\omega$$
$$d)\ \mathsf{Trans}_{\mathcal{U}_\mathcal{L}}(Y, Z) = Z_1 \quad e)\ \mathsf{Trans}_{\mathcal{U}_\mathcal{L}}(Z, Y) = Z_2.$$

Proof. 2 b) and 2 e) are straightforward from Lemma 1 c). 2 c) comes from 2 b) and from the fact that no other final state is accessible from Y.
– 2 d) From Lemma 1 d), $\mathsf{Trans}_{\mathcal{U}_\mathcal{L}}(Y, Z) = \{w \mid w Z_2 \subseteq Y_2\} = Z_1$.
– 1 b) If w is in $\mathsf{Fut}_{\mathcal{U}_\mathcal{L}}(X)$, there exists Y in $\mathfrak{F}_p(\mathcal{L})$ and a factorization of w into xy such that x is in $\mathsf{Trans}_{\mathcal{U}_\mathcal{L}}(X, Y)$ and y is in $\mathsf{Fut}_{\mathcal{U}_\mathcal{L}}(Y)$. Hence $X_1 xy \subseteq X_1 x Y_2^\omega \subseteq \mathcal{L}$ (Lemma 1 b), and $w = xy$ is in X_2. Conversely, if w is in X_2, there exists T in $\mathfrak{F}_p(X_2)$ and a factorization of w into xy, with x in T_1 and y in T_2^ω. Since $X_1 T_1 T_2^\omega \subseteq \mathcal{L}$, there exists Y in $\mathfrak{F}_p(\mathcal{L})$ such that $X_1 T_1 \subseteq Y_1$ and $T_2 \subseteq Y_2$. Therefore $X_1 x Y_2^\omega \subseteq \mathcal{L}$ and y is in Y_2^ω, thus $w = xy$ is in $\mathsf{Trans}_{\mathcal{U}_\mathcal{L}}(X, Y)\mathsf{Fut}_{\mathcal{U}_\mathcal{L}}(Y) \subseteq \mathsf{Fut}_{\mathcal{U}_\mathcal{L}}(X)$.

Let I be the maximal factorization of $\mathcal{L}$ such that $I_2 = \mathcal{L}$. The empty word ε is in I_1 and thus I is an initial state.

– 1 a) Let X be in $\mathfrak{F}(\mathcal{L})$. It holds $X_1 X_2 \subseteq \mathcal{L} = I_2$, hence, by Lemma 1 a), every word of X_1 is in $\mathsf{Past}_{\mathcal{U}_\mathcal{L}}(X)$. Conversely, if w is in $\mathsf{Past}_{\mathcal{U}_\mathcal{L}}(X)$, there exists an initial state X' such that w is in $\mathsf{Trans}_{\mathcal{U}_\mathcal{L}}(X', X)$, thus $X_1' w \subseteq X_1$, and since ε is in X_1', the word w is in X_1.

– 2 a) Let Y be in $\mathfrak{F}_p(\mathcal{L})$. It holds $I_1 Y_1 Y_2^\omega \subseteq I_1 \mathcal{L} = \mathcal{L}$, hence, by Lemma 1 b), every word of Y_1 is in $\mathsf{Past}_{\mathcal{U}_\mathcal{L}}(Y)$. Conversely, if w is in $\mathsf{Past}_{\mathcal{U}_\mathcal{L}}(Y)$, there exists an initial state X such that w is in $\mathsf{Trans}_{\mathcal{U}_\mathcal{L}}(X, Y)$, thus $X_1 w Y_2^\omega \subseteq \mathcal{L}$, and since ε is in X_1, the word w is in $X_1 w \subseteq Y_1$. $\qquad\square$

Proposition 6. *The universal automaton of an ω-language is a finite Büchi automaton in normal form.*

In the universal automaton, the states corresponding to ω-factorizations are transient states, while every pure ω-factorization Y is a final state and elements of $\mathcal{Z}_Y$ are the other states of the final SCC of Y.

From Proposition 5, it is straightforward that:

Proposition 7. *The universal automaton of $\mathcal{L}$ recognizes $\mathcal{L}$.*

The universal automaton of $\mathcal{L}$ is canonical w.r.t. $\mathcal{L}$. We state now that it is indeed universal for $\mathcal{L}$, *i.e.* it contains the morphic image of any Büchi automaton in normal form that recognizes $\mathcal{L}$.

Proposition 8. *(Universality) Let $\mathcal{A}$ be a Büchi automaton in normal form that recognizes $\mathcal{L}$. Then, there exists a morphism φ from $\mathcal{A}$ into $\mathcal{U}_\mathcal{L}$. Moreover, this morphism can be chosen such that transient states of $\mathcal{A}$ map onto transient states of $\mathcal{U}_\mathcal{L}$.*

Depending on the nature of the state, the mapping φ is defined as follows :

1. If p is a transient state, let $X_2 = \{v \mid \mathsf{Past}_\mathcal{A}(p)v \subseteq \mathcal{L}\}$ and $X_1 = \{u \mid uX_2 \subseteq \mathcal{L}\}$. We set $\varphi(p) = (X_1, X_2)$.
2. If p is a final state. Let $Y_1 = \{u \mid u\mathsf{Trans}_\mathcal{A}(p,p)^\omega \subseteq \mathcal{L}\}$ and $Y_2 \in max\{T \mid Y_1 T^\omega \subseteq \mathcal{L}$ and $\mathsf{Trans}_\mathcal{A}(p,p) \subseteq T\}$. We set $\varphi(p) = (Y_1, Y_2)$.
3. If p belongs to a SCC containing a final state q distinct from p. Let $Y = \varphi(q)$, $Z_1 = \{u \mid u\mathsf{Trans}_\mathcal{A}(p,q) \subseteq Y_2\}$ and $Z_2 = \{v \mid Z_1 v \subseteq Y_2\}$. If $(Z_1, Z_2) = (Y_2, Y_2)$, then let $K = Y$, otherwise $K = Z$. We set $\varphi(p) = K$.

Every automaton in normal form that recognizes the ω-language $\mathcal{L}$ and that fulfils the universal property contains the universal automaton. The maximality of factorizations implies that the merging of distinct states leads to accept more ω-words.

Proposition 9. *Let $\mathcal{V}$ be an automaton in normal form recognizing $\mathcal{L}$ such that there exists a morphism φ from $\mathcal{U}_\mathcal{L}$ into $\mathcal{V}$. Then φ is injective.*

Propositions 7,8 and 9 put together give the main result of this paper.

Theorem 3. *For every ω-rational language $\mathcal{L}$, the universal automaton of $\mathcal{L}$ is the smallest Büchi automaton in normal form that recognizes $\mathcal{L}$ in which every equivalent Büchi automaton in normal form has a morphic image.*

5 Conclusion

This paper introduces factorizations of ω-rational languages. They lead to the definition of the universal automaton of such a language. This automaton is effectively computable, since the maximal ω-factorizations are computable either from a finite ω-semigroup that recognizes the language or from a prophetic automaton. The maximal pure ω-factorizations are also computable from the same ω-semigroup, but it remains open whether they can be efficiently computed from some automaton accepting the language.

Like for rational languages on finite words, the universal automaton may be useful, in spite of its size, which is at most exponential in the size of the syntactic ω-semigroup, in the proofs of existence of Büchi automata with specific properties. On finite words, the universal automaton allows for instance to prove that a reversible rational language can be recognized by a NFA which is both reversible and star-height minimal [4].

Moreover, since every automaton has a morphic image in the universal automaton, it can be a tool for the construction of automata with a small number of states. In the case of Büchi automata, since the universal automaton is in normal form, the computation of a Büchi automaton with a minimal number of states is not as straightforward as in the case of NFA. It is probably a question which deserves more studies.

References

1. CARTON, O., AND MICHEL, M. Unambiguous Büchi automata. *Theoret. Comput. Sci. 297* (2003), 37–81.
2. CONWAY, J. H. *Regular algebra and finite machines*. Mathematics series. Chapmann and Hall, London, 1971.
3. LOMBARDY, S. On the construction of reversible automata for reversible languages. In *ICALP 2002* (2002), vol. 2380 of *Lect. Notes Comp. Sci.*, Springer, pp. 170–182.
4. LOMBARDY, S., AND SAKAROVITCH, J. Star height of reversible languages and universal automata. In *LATIN 2002* (2002), vol. 2286 of *Lect. Notes Comp. Sci.*, Springer, pp. 76–90.
5. LOMBARDY, S., AND SAKAROVITCH, J. The universal automaton. In *Logic and Automata*, vol. 2 of *Texts in Logic and Games*. Amsterdam University Press, 2008, pp. 457–504.
6. PERRIN, D., AND PIN, J.-É. Semigroups and automata on infinite words. In *Semigroups, Formal Languages and Groups* (1995), Kluwer academic publishers, pp. 49–72.
7. PERRIN, D., AND PIN, J.-É. *Infinite Words*, vol. 141 of *Pure and Applied Mathematics*. Elsevier, 2004.
8. POLÁK, L. Minimalizations of nfa using the universal automaton. *Int. J. Found. Comput. Sci. 16*, 5 (2005), 999–1010.
9. SAKAROVITCH, J. *Elements of Automata Theory*. Cambridge University Press, 2009.