Ultimately Periodic Words of Rational ω-Languages

Hugues Calbrix * Maurice Nivat * Andreas Podelski †

Abstract

In this paper we initiate the following program: Associate sets of finite words to Büchi-recognizable sets of infinite words, and reduce algorithmic problems on Büchi automata to simpler ones on automata on finite words. We know that the set of ultimately periodic words $UP(L)$ of a rational language of infinite words L is sufficient to characterize L, since $UP(L_1) = UP(L_2)$ implies $L_1 = L_2$. We can use this fact as a test, for example, of the equivalence of two given Büchi automata. The main technical result in this paper is the construction of an automaton which recognizes the set of all finite words $u \cdot \$ \cdot v$ which naturally represent the ultimately periodic words of the form $u \cdot v^\omega$ in the language of infinite words recognized by a given Büchi automaton.

1 Introduction

Büchi automata recognizing sets of infinite words appear as a major tool in modelizing the behavior of a number of computing systems including distributed and real-time systems and circuits. The standard theoretical results about the decidability of the equivalence of two Büchi automata do not lead to efficient algorithms for equality test or optimisation of such automata, see e.g. Safra[5] or Sistla, Vardi and Volper [6] (a question about which almost nothing is known). The basic idea underlying the present paper is that a set of infinite words recognized by a Büchi automaton is entirely known when we know the subset of ultimately periodic words (of the form $u \cdot v^\omega$) it contains, and we prove that this set is finitely representable since the set of finite words $u \cdot \$ \cdot v$ corresponding to all the $u \cdot v^\omega$ is rational, i.e. recognizable by a finite automaton. This fact brings the hope that a number of constructions which are presently outwardly performed on Büchi automata can be performed on simple dfa's. This is already the case for the S1S logic (see [7]) for which this method brings an described in [2].

Two main theorems are proved in this paper. The first one states the rationality of $L_\$$, the language of finite representations of ultimately peri-

*LITP, Université de Paris 7, 2,place Jussieu ,75251 Paris Cedex 05, France. e-mail: {calbrix,tcsmn}@litp.ibp.fr

†Digital PRL, 85, avenue Victor Hugo, 92563 Rueil-Malmaison, France. e-mail: podelski@prl.dec.com

odic words of an arbitrary rational ω-language L, and its proof brings a construction of an automaton that recognizes $L_\$$. The second one states a nice characterization of the languages K that are $L_\$$ for a given rational ω-language L and also brings a construction of a Büchi automaton recognizing L.

Section 3 describes informally the first construction and two representative examples are shown. The formal proof of the first result lies in Section 4. In Section 5, we study the determinisation of the previous construction and we give an upper bound of its number of states. Section 6 is devoted to the proof of the second theorem and also gives a bound to the number of states of the second construction. In Section 7, we raise some questions about rational languages contained in $A^* \cdot \$ \cdot A^+$ and the set of ultimately periodic words that they represent. Section 8 concludes this paper.

2 Basic Definitions

Let A be a finite set called the *alphabet*. We denote A^* the set of *finite words* on A — finite sequence of elements of A. We note ϵ the empty sequence, which is called the *empty word*. We denote A^+ the set of non-empty words, i.e. $A^+ = A^* \setminus \{\epsilon\}$. Let u be a finite word. We denote by $|u|$ the length of the sequence u. The length of the empty word ϵ is thus 0. We denote by A^ω the set of *infinite words* on A — infinite sequences of elements of A. A *language* is a subset M of A^*, and ω-*language* a subset L of A^ω.

A *finite automaton* $\mathcal{A}$ is a tuple (Q, I, D, E) made of a finite set Q, the elements of which are the *states* of the automaton, a subset I of Q of *initial* states, a subset D of Q of *distinguished* states, and a subset E of $Q \times A \times Q$, the elements of which are the *edges* of the automaton. It will be convenient to number the elements of Q. We will then write $Q = \{q_1, \dots q_m\}$.

Let $u = u(1) \cdot \ldots \cdot u(k)$ be a finite word. A word $c = c(0) \cdot \ldots \cdot c(k+1)$ of Q^+ is a calculus of $\mathcal{A}$ on u if $(c(i), u(i), c(i+1)) \in E$ for each i such that $1 \leq i \leq k$. This calculus is *successful* if $c(0) \in I$ and $c(k+1) \in D$. We denote $L(\mathcal{A})$ the language of finite words u such that there is a successful calculus of $\mathcal{A}$ on u. In this case, the elements of D are called *final* states and D is denoted F. The set of languages $L(\mathcal{A})$ for some automaton $\mathcal{A}$, is denoted $\mathcal{R}at(A^*)$, and its elements are called *rational languages*.

Let $\mathcal{A}$ be a finite automaton and $v \in A^+$ a non-empty finite word. When there exists a calculus $c \in Q^+$ of $\mathcal{A}$ on v such that $c = p \cdot c' \cdot q$, we will write $p \xrightarrow[\mathcal{A}]{v} q$. When in addition $c' \cdot q$ contains a distinguished state, we will write $p \overset{v}{\underset{\mathcal{A}}{-\!\circ\!\!\rightarrow}} q$, and on the other hand, when $c' \cdot q$ contains no distinguished state, we will write $p \overset{v}{\underset{\mathcal{A}}{\rightsquigarrow}} q$.

Let $\alpha = \alpha(0) \cdot \alpha(1) \cdot \ldots$ be an infinite word. A word $\chi = \chi(0) \cdot \chi(1) \cdot \ldots$ of Q^ω is a *calculus* of $\mathcal{A}$ on α if $(\chi(i), \alpha(i), \chi(i+1)) \in E$ for each integer i. This calculus is *successful* if $\chi(0) \in I$ and if there exists a distinguished state q of D such that $\chi(k) = q$ for infinitely many integers k. We denote $L^\omega(\mathcal{A})$ the ω-language of infinite words α such that there is a successful calculus of $\mathcal{A}$ on α. In this case, the elements of D are called *repeated* states and D is denoted R and $\mathcal{A}$ is called a *Büchi automaton*. The set of ω-languages $L^\omega(\mathcal{A})$ for some automaton $\mathcal{A}$ is denoted $\mathcal{R}at(A^\omega)$ and its elements are called *rational* ω-languages.

We denote $UP(A^\omega)$ the set $\{u \cdot v^\omega \mid (u, v) \in A^* \times A^+\}$, the elements of which are the *ultimately periodic* words. Let L be an ω-language, we denote $UP(L)$ the set $L \cap UP(A^\omega)$ of all ultimately periodic words of L. Let α be an ultimaltely periodic word of A^ω. A word $v \in A^+$ is a *period* of α if there exists a word $u \in A^*$ such that $\alpha = u \cdot v^\omega$. Similarly, a word $u \in A^*$ is a *prefix* of α if there is a period v of α such that $\alpha = u \cdot v^\omega$. This definition of a prefix is thus more restrictive than the usual one. Indeed, a isn't a prefix of $aa \cdot b^\omega$, for there is no word $v \in A^+$ such that $aa \cdot b^\omega = a \cdot v^\omega$.

Fact 1 *Let L_1 and L_2 be two rational ω-languages such that $UP(L_1) = UP(L_2)$, then $L_1 = L_2$.*

Proof The ω-language $(L_1 \cup L_2) \setminus (L_1 \cap L_2)$ does not contain any ultimately periodic word and it is a rational ω-language, because the set $\mathcal{R}at(A^\omega)$ is closed under boolean combinations. However, every non-empty rational ω-language contains at least one ultimately periodic word. Thus $(L_1 \cup L_2) \setminus (L_1 \cap L_2)$ is the empty set and $L_1 = L_2$. $\diamondsuit$

The set of ultimately periodic words of a rational ω-language is thus characteristic of this ω-language. The ultimately periodic word $u \cdot v^\omega$ on the alphabet A may be represented by the finite word $u \cdot \$ \cdot v$ on the alphabet $A \cup \$$, where $\$$ is a dummy symbol which is not already in A. Let L be a rational ω-language. We define the language $L_\$ = \{u \cdot \$ \cdot v \mid u \cdot v^\omega \in L\}$ on the alphabet $A \cup \$$, to be the set of all the finite words which represent ultimately periodic words of L. The Fact 1 allows us to say that $L_\$$ characterizes the rational ω-language L.

3 Finite Words

Let L be a rational ω-language and $\mathcal{A} = (Q, I, R, E)$ a Büchi automaton which recognizes it (we set $Q = \{q_1, \ldots, q_m\}$.) For each r such that $1 \leq r \leq m$, we set $M_r = \{u \in A^* \mid \exists q \in I, q \xrightarrow{u} q_r\} = L(Q, I, \{q_r\}, E)$ and $N_r = \{v \in A^+ \mid v^\omega \in L^\omega(Q, \{q_r\}, R, E)\}$. It is clear that for each pair of words $(u, v) \in M_r \times N_r$, $u \cdot v^\omega \in L$, because a successful calculus of $\mathcal{A}$ on $u \cdot v^\omega$ may be built from a calculus of $(Q, I, \{q_r\}, E)$ on u leading to q_r and a successful calculus of $(Q, \{q_r\}, R, E)$ on v^ω. Moreover, for each ultimately periodic word $u \cdot v^\omega \in L$, there exists a $q_r \in Q$ such that $u \in M_r$ and $v \in N_r$. This q_r is the state reached after the reading of u in a successful calculus of $\mathcal{A}$ on $u \cdot v^\omega$. We may decompose the previously defined language $L_\$$, using the languages M_r and N_r in the following way.

$$L_\$ = \bigcup_{r=0}^{m} M_r \cdot \$ \cdot N_r \qquad (1)$$

Languages M_r are made of prefixes of ultimately periodic words of L and these languages are rational, because they are recognized by automata $(Q, I, \{q_r\}, E)$. Languages N_r are made of periods of ultimately periodic words of L. We will build automata which recognize languages N_r to show that they are rational too. The rationality of $L_\$$ will follow from this fact.

It might be noticed that there are various ways to show the rationality of $L_\$$. We can show that the syntactic congruence of $L_\$$ and Arnold's congruence of L (defined in [1]) are the same on the set A^+ (see [3]). Then the syntactic congruence of $L_\$$ is of finite index and $L_\$$ is thus rational. It is also possible to use the equivalence between S1S-logical definability and rationality for ω-languages (see e.g. [7]) to construct an automaton recognizing $L_\$$ from a logical formula defining L (this procedure is described in [2]). However, the direct construction of an automaton is the most efficient way to produce a recognizing device for $L_\$$.

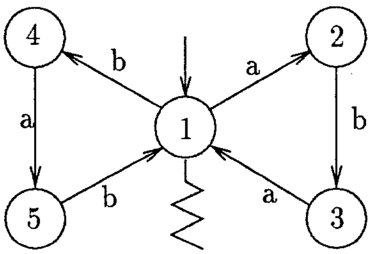

figure 1

A word v^ω is recognized by the automaton $(Q, \{q_r\}, R, E)$ if there is a successful calculus of this automaton on v^ω. This calculus runs along one or several loops—i.e. cyclic sequences of states—which contains repeated states of R. For exemple, let $L = (aba + bab)^\omega$ be the language recognized by the automaton of the *Figure 1*. The word $(aba)^\omega$ is recognized, and the infinite sequence of states $(123)^\omega$ is a successful calculus of the automaton on $(aba)^\omega$. This calculus defines a loop—1231—which runs through a repeated state—1.

The word aba is in the language N_1 of periods of ultimately periodic words of L recognized from the state 1, and we just need to know the calculus of the automaton on the word aba to find the loop 1231. This is not always the case, as the next example will show it. The word $(ab)^\omega$ is recognized too, the infinite sequence of states $(123145)^\omega$ is a successful calculus of the automaton on this word and the loop found here is the sequence 1231451. The word ab is member of the language N_1 but the calculus of the automaton on the word ab doesn't permit us to find the loop, which only appears in the calculus of $(ab)^3$.

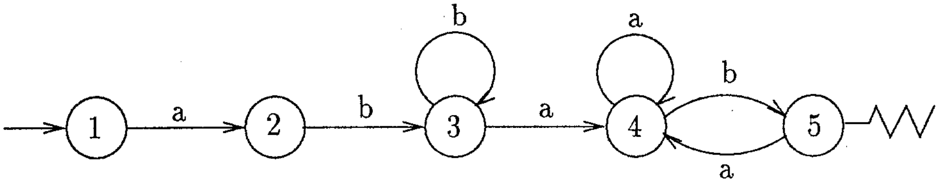

figure 2

Another example is given by the language $L = ab^+ \cdot (a^+b)^\omega$, recognized by the automaton of the *Figure 2*. The word $(ab)^\omega$ is recognized by this automaton and the infinite sequence of states $123(45)^\omega$ is a successful calculus of the automaton on the word $(ab)^\omega$. The word ab is element of the language N_1 of the periods of L which can be read from the state 1 of the automaton, but a calculus on $ababab$ is necessary to find a loop—here 545. It becomes clear with this example that the first state of the calculus isn't necessarily involved in the loop found in this calculus.

The principle of the construction that we are going to describe is to simulate calculi of the automaton $\mathcal{A}$ which recognize the language L, starting from each state of the automaton $\mathcal{A}$. This simulation leads to a vector-state which contains as components ends of simulated calculi with an element of the set $\{0, 1\}$ which is 1 if and only if the simulated calculus contains a repeated state. Final states are those from which a loop of $\mathcal{A}$ containing a repeated state can be built. For the first example, the calculus of the automaton recognizing N_1 on the word ab may be—the state denoted $\square$ is added to $\mathcal{A}$ to make it complete

$$\begin{pmatrix} 1,0 \\ 2,0 \\ 3,0 \\ 4,0 \\ 5,0 \end{pmatrix} \xrightarrow{a} \begin{pmatrix} 2,0 \\ \square \\ 1,1 \\ 5,0 \\ \square \end{pmatrix} \xrightarrow{b} \begin{pmatrix} 3,0 \\ \square \\ 4,1 \\ 1,1 \\ \square \end{pmatrix}$$

From this calculus, we can build the following calculi of the automaton $\mathcal{A}$ on the word ab,

$$1 \overset{ab}{\underset{\mathcal{A}}{\rightsquigarrow}} 3,\ 3 \overset{ab}{\underset{\mathcal{A}}{-\!\circ\!\!\rightarrow}} 4 \text{ and } 4 \overset{ab}{\underset{\mathcal{A}}{-\!\circ\!\!\rightarrow}} 1,$$

which permit us to find a loop containing a repeated state in a calculus of $\mathcal{A}$ on the word $(ab)^\omega$. Moreover, we can make the calculus begin with the state 1, 3 or 4, which shows that the word ab is element of N_1, N_3 and N_4.

4 First Construction

Formally, let L be an ω-language recognized by an automaton $\mathcal{A} = (Q, I, R, E)$ such that $Q = \{q_1, \ldots, q_m\}$. We suppose without loss of generality that the automaton $\mathcal{A}$ is complete, i.e. that $\{q \mid (p, a, q) \in E\} \neq \emptyset$ for each pair $(q, a) \in Q \times A$. For each state q_r of the automaton $\mathcal{A}$, we will build an automaton $\mathcal{A}_{N_r}$ which recognizes the language N_r previously defined.

The automaton $\mathcal{A}_{N_r}$ is built from the set of states $(Q \times \{0,1\})^{|Q|}$. The initial state is the vector-state $\vec{q_0} = ((q_1, 0), \ldots, (q_m, 0))$. The tuple $(\vec{p}, a, \vec{p}')$— with $\vec{p} = ((p_1, f_1), \ldots, (p_m, f_m))$ and $\vec{p}' = ((p'_1, f'_1), \ldots, (p'_m, f'_m))$— is an edge of $\mathcal{A}_{N_r}$ if, for each i such that $1 \leq i \leq m$, $(p_i, a, p'_i) \in E$ and if $p'_i \in R$ then $f'_i = 1$ else $f'_i = f_i$. A state $\vec{p} = ((p_1, f_1), \ldots, (p_m, f_m))$ is a final state if the following condition is verified. Let $(i_k)_{1 \leq k \leq m}$ be the finite sequence of integers defined by the relation $p_k = q_{i_k}$, and $(j_k)_{k \geq 0}$ be the infinite sequence defined recursively by $j_0 = r$ and $j_{k+1} = i_{j_k}$ for each $k \geq 0$. This sequence ranges only over a finite set of values. Let thus s be the smallest integer satisfying $j_s \in \{j_k \mid 0 \leq k < s\}$ and s' the only integer such that $s' < s$ and $j_{s'} = j_s$. Then, the state $\vec{p}$ is final if and only if $1 \in \{f_{j_k} \mid s' \leq k \leq s\}$.

The following lemma states the fundamental property of the automaton $\mathcal{A}_{N_r}$. A calculus of $\mathcal{A}_{N_r}$ begining with $\vec{q_0}$ on a word v contains for each state q_i a calculus of the automaton $\mathcal{A}$ begining with q_i on the word v. Moreover, a state of the automaton $\mathcal{A}_{N_r}$ reached by the reading of v from the state $\vec{q_0}$ can be built from the calculi of $\mathcal{A}$ on v starting from each state of $\mathcal{A}$.

Lemma 2 *Let $v \in A^*$ and $\vec{p} = ((p_1, f_1), \ldots, (p_m, f_m))$ be a state of $\mathcal{A}_{N_r}$. Then $\vec{q_0} \xrightarrow[\mathcal{A}_{N_r}]{v} \vec{p}$ if and only if, for each i such that $1 \leq i \leq m$, $q_i \overset{v}{\underset{\mathcal{A}}{\rightsquigarrow}} p_i$ if $f_i = 0$ and $q_i \overset{v}{\underset{\mathcal{A}}{-\!\circ\!\!\rightarrow}} p_i$ if $f_i = 1$.*

Proof Let $v \in A^*$ and $\vec{p}$ a state of $\mathcal{A}_{N_r}$. We will show the lemma with an induction on the length of the word v. If $v = \epsilon$, then the lemma trivially is true. Thus, we assume that $v \neq \epsilon$ and we set $v = u \cdot a$ with $u \in A^*$ and $a \in A$.

Let us assume that $\vec{q_0} \xrightarrow[\mathcal{A}_{N_r}]{v} \vec{p}$. Let $\vec{p}\,'$ a state of $\mathcal{A}_{N_r}$ such that $\vec{q_0} \xrightarrow[\mathcal{A}_{N_r}]{u} \vec{p}\,'$ and $\vec{p}\,' \xrightarrow[\mathcal{A}_{N_r}]{a} \vec{p}$. We set $\vec{p}\,' = ((p'_1, f'_1), \ldots, (p'_m, f'_m))$. We deduce from the induction hypothesis, that $q_i \xrightarrow[\mathcal{A}]{u} p'_i$ for each i such that $1 \leq i \leq m$. Since $(\vec{p}\,', a, \vec{p})$ is an edge of $\mathcal{A}_{N_r}$, we deduce from the definition of $\mathcal{A}_{N_r}$ that (p'_i, a, p_i) is an edge of $\mathcal{A}$, and that $q_i \xrightarrow[\mathcal{A}]{v} p_i$, for each i such that $1 \leq i \leq m$. Moreover, if $f_i = 1$ then $f'_i = 1$ and in this case $q_i \overset{u}{\underset{\mathcal{A}}{\multimap}} p'_i$, or $f'_i = 0$ and in that case p_i is a repeated state. In both cases, we get that $q_i \overset{v}{\underset{\mathcal{A}}{\multimap}} p_i$. To conclude, if $f_i = 0$ then $f'_i = 0$, and we deduce from induction hypothesis that $q_i \overset{u}{\underset{\mathcal{A}}{\leadsto}} p'_i$. The state p_i isn't a repeated state and thus $q_i \overset{v}{\underset{\mathcal{A}}{\leadsto}} p_i$.

Let us now assume that $q_i \overset{v}{\underset{\mathcal{A}}{\leadsto}} p_i$ for each i such that $f_i = 0$ and that $q_i \overset{v}{\underset{\mathcal{A}}{\multimap}} p_i$ for each i such that $f_i = 1$. For each i such that $1 \leq i \leq m$, there is a state p'_i from Q such that $q_i \xrightarrow[\mathcal{A}]{u} p'_i$ and (p'_i, a, p_i) is an edge of $\mathcal{A}$. Moreover, if $f_i = 1$ and p_i isn't a repeated state, then we can choose p'_i such that $q_i \overset{u}{\underset{\mathcal{A}}{\multimap}} p'_i$, and we set then $f'_i = 1$. If $f_i = 1$ and p_i is a repeated state, we set $f'_i = 0$ if $q_i \overset{u}{\underset{\mathcal{A}}{\leadsto}} p'_i$ and $f'_i = 1$ in the other case. If $f_i = 0$, we can choose p'_i such that $q_i \overset{u}{\underset{\mathcal{A}}{\leadsto}} p'_i$, and we simply set $f'_i = 0$. We deduce from the induction hypothesis that the state $\vec{p}\,' = ((p'_1, f'_1), \ldots, (p'_m, f'_m))$ thus defined is such that $\vec{q_0} \xrightarrow[\mathcal{A}_{N_r}]{u} \vec{p}\,'$. Then, we see from the definition of the automaton $\mathcal{A}_{N_r}$ that $(\vec{p}\,', a, \vec{p})$ is an edge of $\mathcal{A}_{N_r}$. Thus, we conclude that $\vec{q_0} \xrightarrow[\mathcal{A}_{N_r}]{v} \vec{p}$. $\diamondsuit$

It remains to show the equality between N_r and $L(\mathcal{A}_{N_r})$. So, let v be a word of $L(\mathcal{A}_{N_r})$, let $\vec{p} = ((p_1, f_1), \ldots, (p_m, f_m))$ be a final state of $\mathcal{A}_{N_r}$ such that $\vec{q_0} \xrightarrow[\mathcal{A}_{N_r}]{v} \vec{p}$, and let sequences $(i_k)_{1 \leq k \leq m}$ and $(j_k)_{k \geq 0}$ and integers s and s' be defined as previously. From the previous lemma, we deduce the existence of calculi $b_1, \ldots, b_m$ of the automaton $\mathcal{A}$ on the word v such that $b_k = q_k \cdot b'_k \cdot p_k$ for each k such that $1 \leq k \leq m$ (we set $c_k = q_k \cdot b'_k$.) From the definitions of sequences $(i_k)_{1 \leq k \leq m}$ and $(j_k)_{k \geq 0}$, we deduce the equalities $p_{j_k} = q_{i_{j_k}} = q_{j_{k+1}}$. The infinite word $c_{j_0} \cdot \ldots \cdot c_{j_{s'-1}} \cdot (c_{j_{s'}} \cdot \ldots \cdot c_{j_{s-1}})^\omega$ of Q^ω is thus a calculus of $\mathcal{A}$ on the infinite word v^ω. Moreover, $\vec{p}$ is a final state of $\mathcal{A}_{N_r}$, then $f_{j_k} = 1$ for an integer k such that $s' \leq k < s$, and we deduce from the Lemma 2 that $c'_{j_k} \cdot q_{j_{k+1}}$ contains a repeated state. Because the first state of the previous infinite calculus is $q_{j_0} = q_r$, this is a successful calculus of the automaton $(Q, \{q_r\}, R, E)$ on the infinite word v^ω. The inclusion $L(\mathcal{A}_{N_r}) \subseteq N_r$ is thus proved.

Conversely, let v be a word of N_r. If $\mathcal{A}$ isn't deterministic, there exist non-regular calculus of $\mathcal{A}$ on v^ω. However, we will show in the next lemma the existence of a particular ultimately periodic calculus of $\mathcal{A}$ on v^ω which can be used to build a successful calculus of $\mathcal{A}_{N_r}$ on v.

Lemma 3 *Let $v \in N_r$—i.e. such that $v^\omega \in L^\omega(Q, \{q_r\}, R, E)$. Then, there is a successful calculus $\pi \in Q^\omega$ of the automaton $(Q, \{q_r\}, R, E)$ on v^ω which satisfy the following property. There is two integers s and s', $s' < s$, and some words $c_0, \ldots, c_{s-1} \in Q^*$ such that $|c_k| = |v|$, $c_k = p_k \cdot c'_k$ with $p_k \in Q$ and $p_k \neq p_l$ for each pair of integers k, l such that $0 \leq k, l < s$ and $k \neq l$ and these words verify $\pi = c_0 \cdot \ldots \cdot c_{s'_1} \cdot (c_{s'} \cdot \ldots \cdot c_{s-1})^\omega$.*

Proof Let $v \in N_r$ and $\chi = c_0 \cdot c_1 \cdots$ be a successful calculus of $(Q, \{q_r\}, R, E)$ on v^ω with $|c_k| = |v|$ for each integer $k \geq 0$. This calculus isn't necessarily ultimately periodic because $\mathcal{A}$ can be non-deterministic. For each integer $k \geq 0$, we set $c_k = p_k \cdot c'_k$, with $p_k \in Q$. Let then s be the least integer satisfying $p_s \in \{p_k \mid 0 \leq k < s\}$—such an integer exists because Q is a finite set—and s' be the integer such that $s' < s$ and $p_{s'} = p_s$. There are two possibilities. In the first case, the word $c_{s'} \cdots c_{s-1}$ contains a repeated state and then $c_0 \cdots c_{s'_1} \cdot (c_{s'} \cdots c_{s-1})^\omega$ is a successful calculus of $(Q, \{q_r\}, R, E)$ on v^ω which satisfy hypothesis of the lemma. In the other case, $\chi' = c_0 \cdot \ldots \cdot c_{s'-1} \cdot c_s \cdot \ldots$ is a successful calculus of $(Q, \{q_r\}, R, E)$ on v^ω. Then we repeat the whole process with χ' until we are in the first case. Because we're removing a non-empty factor of χ at each step and χ is successful, we are sure that the process will stop in a finite number of steps. $\diamondsuit$

Then, let π be a calculus of $\mathcal{A}$ on v^ω satisfying the hypothesis of Lemma 3. We can set $q_1 = p_0, \ldots, q_s = p_{s-1}$ without loss of generality, the other states of Q are numbered arbitrarily, and we also set $p_s = p_{s'}$. For all k such that $1 \leq k \leq s$, the word $c_{k-1} \cdot p_k$ is a calculus of $\mathcal{A}$ on v, and thus $q_k \xrightarrow[\mathcal{A}]{v} p_k$. On the other hand, $\mathcal{A}$ is complete, and then for all k such that $s < k \leq m$, there is a state p_k such that $q_k \xrightarrow[\mathcal{A}]{v} p_k$. For each k such that $1 \leq k \leq s$, we set $f_k = 1$ if $c'_{k-1} \cdot p_k$ contains a repeated state, $f_k = 0$ otherwise. For each k such that $s < k \leq m$, we set $f_k = 0$ if $q_k \overset{v}{\underset{\mathcal{A}}{\rightsquigarrow}} p_k$, and $f_k = 1$ otherwise. The state $\vec{p} = ((p_1, f_1), \ldots, (p_m, f_m))$ thus defined is such that $\vec{q}_0 \xrightarrow[\mathcal{A}_{N_r}]{v} \vec{p}$. For this state, the sequence $(j_k)_{k \geq 0}$ is defined by $j_0 = 1$, $j_k = k+1$ for all $k < s$, and $j_k = j_{k-(s-s')}$ for all $k \geq s$. s is the least integer verifying $j_s \in \{j_k \mid 1 \leq k < s\}$ and s' is such that $s' < s$ and $j_{s'} = j_s$. Moreover, π is successful then $f_{j_k} = 1$ for an integer k such that $s' \leq k \leq s$. This shows that $\vec{p}$ is a final state of $\mathcal{A}_{N_r}$ and finishes the proof of the inclusion $N_r \subseteq L(\mathcal{A}_{N_r})$.

The languages N_r are recognized by the automata $\mathcal{A}_{N_r}$ and are thus rational. From the equality (1), we deduce that the language $L_\$$ is rational too. Finally, we have shown the following proposition.

Proposition 4 *Let L be a rational ω-language on the alphabet A and let $L_\$$ be the language of finite words on the alphabet $A \cup \$$, defined by $L_\$ = \{u \cdot \$ \cdot v \mid u \cdot v^\omega \in L\}$. Then $L_\$$ is rational.* $\diamondsuit$

It is easy to construct an automaton recognizing $L_\$$ from the automata $\mathcal{A}$ and $\mathcal{A}_{N_r}$. Indeed, let $\mathcal{A}_\$$ be the disjoint union of automata $\mathcal{A}$ and $\mathcal{A}_{N_r}$, for each r, to which we're adding the edges $(q_r, \$, \vec{q_0}_r)$—$\vec{q_0}_r$ is the initial state of $\mathcal{A}_{N_r}$. The initial states of $\mathcal{A}_\$$ are those of $\mathcal{A}$ and the final states of $\mathcal{A}_\$$ are those of all $\mathcal{A}_{N_r}$. Then obviously $L(\mathcal{A}_\$) = L_\$$.

5 Determinising $\mathcal{A}_\$$

The automaton $\mathcal{A}_\$$ that we built in the previous paragraph isn't deterministic. One reason for this is that it contains $\mathcal{A}$, which itself is not generally deterministic. However, accessible states of the subset automata built from $\mathcal{A}_\$$ have a particular shape, which provides a simple representation of these states and a bound to its number.

We first build for each state q_r of $\mathcal{A}$ the subset automaton $\mathcal{P}(\mathcal{A}_{N_r})$ of the automaton $\mathcal{A}_{N_r}$. Its initial state is the singleton $\{\vec{q_0}\}$, and we denote δ its transition function. Let P be an accessible state of $\mathcal{P}(\mathcal{A}_{N_r})$ and v a word of A^* such that $\delta(\{\vec{q_0}\}, v) = P$. Let $\vec{p}$ and $\vec{p}'$ two states of $\mathcal{A}_{N_r}$, members of P — $\vec{p} = ((p_1, f_1), \ldots, (p_m, f_m))$, $\vec{p}' = ((p'_1, f'_1), \ldots, (p'_m, f'_m))$. Then, each state $\vec{p}'' = ((p''_1, f''_1), \ldots, (p''_m, f''_m))$ such that $(p''_k, f''_k) \in \{(p_k, f_k), (p'_k, f'_k)\}$ for each k such that $1 \leq k \leq m$ is a member of P. This is a direct consequence of Lemma 2. The state P is thus entirely defined by the sets $P_k = \{(p_k, f_k) \mid \vec{p} = ((p_1, f_1), \ldots, (p_m, f_m)) \in P\}$, i.e. P is the set of states $\vec{p} = ((p_1, f_1), \ldots, (p_m, f_m))$ such that $(p_k, f_k) \in P_k$ for each k such that $1 \leq k \leq m$. The set of states of the subset automaton is in bijective correspondence with the set $(\mathcal{P}(Q \times \{0, 1\}))^m$, which contains 2^{2m^2} elements.

The automata $\mathcal{A}_{N_r}$, and thus $\mathcal{P}(\mathcal{A}_{N_r})$, have the same stucture — the only thing that changes is final states — and there is a straightfoward construction to build a deterministic automaton recognizing a language such as $N = \cup_{i=1}^{k} N_{r_i}$, the union of languages N_r. This automaton is isomorphic to the common structure of $\mathcal{P}(\mathcal{A}_{N_r})$, and its set of final states is the union of the final states of automata $\mathcal{P}(\mathcal{A}_{N_{r_i}})$.

Now we build a deterministic automaton that recognizes $L_\$$. This automaton is the disjoint union of $\mathcal{P}(\mathcal{A})$, the subset automaton of the automaton $\mathcal{A}$, and of automata that we built previously recognizing each language N, the union of the languages N_r to which we add edges $(P, \$, \vec{q}_{0_P})$, where $\vec{q}_{0_P}$ is the initial state of the automaton recognizing the language $\cup_{r \in P} N_r$. The automaton we have built is deterministic and recognizes the language $L_\$$. There are at most 2^m states in $\mathcal{P}(\mathcal{A})$, and there are at most 2^m unions of languages N_r, which are recognized by automata with at most 2^{2m^2} states. Finally, there are at most $2^m + 2^{2m^2+m}$ states in this automaton.

6 Infinite Words and Second Construction

Let L be a rational ω-language and $L_\$$ the rational language defined in the previous paragraphs. Let $u \cdot \$ \cdot v$ be a word in $L_\$$ and $u' \cdot \$ \cdot v'$ a word in $A^* \cdot \$ \cdot A^+$ such that $u \cdot v^\omega = u' \cdot v'^\omega$. It is then clear that $u' \cdot \$ \cdot v'$ is an element of $L_\$$. Let us define the equivalence relation $\overset{UP}{\equiv}$ on the language $A^* \cdot \$ \cdot A^+$ in the following way.

$$u \cdot \$ \cdot v \overset{UP}{\equiv} u' \cdot \$ \cdot v' \text{ if and only if } u \cdot v^\omega = u' \cdot v'^\omega,$$

Then, $L_\$$ is saturated by $\overset{UP}{\equiv}$.

Let K be a rational language of $(A \cup \$)^*$ contained in $A^* \cdot \$ \cdot A^+$. A necessary condition for K to be $L_\$$ for a rational ω-language L is that K is saturated by $\overset{UP}{\equiv}$. We will show that this condition is sufficient too, and we will construct an automaton that recognizes L. We first need the following lemma.

Lemma 5 *Let M and N be two languages of A^* such that $M \cdot N^* = M$ and $N^+ = N$. Then, for each infinite word $\alpha \in A^\omega$, $\alpha \in UP(M \cdot N^\omega)$ if and only if there exist two words $u \in M$ and $v \in N$ such that $u \cdot v^\omega = \alpha$.*

Proof It is clear that for each words $u \in M$ and $v \in N$, $u \cdot v^\omega \in M \cdot N^\omega$. Conversely, let $\alpha = u \cdot v^\omega$ be a ultimately periodic word of $M \cdot N^\omega$, and $u_0, u_1, \ldots$ a sequence of words such that $u_0 \in M$, $u_i \in N$ for each $i > 0$ and $u_0 \cdot u_1 \cdot \ldots = u \cdot v^\omega$. We set $l = |v|$, $l_i = |u_0 \cdot \ldots \cdot u_i|$ for each integer i, and $P = \{l_i \mid i \in \mathrm{N}\}$. P is an infinite subset of N, thus there is an integer k such that $P \cap (l\mathrm{N} + k)$ is infinite. Let n_1 and n_2 be two integers such that $0 < n_1 < n_2$, $l_{n_1} > |u|$, and l_{n_j} is in $l\mathrm{N} + k$ for $j = 1$ and 2. We can then find two words v_1 and v_2 such that $v = v_1 \cdot v_2$, and two integers k_1 and k_2 such that $u_0 \cdot \ldots \cdot u_{n_1} = u \cdot v^{k_1} \cdot v_1$ and $u_{n_1+1} \cdot \ldots \cdot u_{n_2} = v_2 \cdot v^{k_2} \cdot v_1$. The two ultimately periodic words $u \cdot v^\omega$ and $u_0 \cdot \ldots \cdot u_{n_1} \cdot (u_{n_1+1} \cdot \ldots \cdot u_{n_2})^\omega$ are equal. We deduce from the hypothesis on languages M and N that $u_0 \cdot \ldots \cdot u_{n_1} \in M$ and $u_{n_1+1} \cdot \ldots \cdot u_{n_2} \in N$, and this ends the proof. $\diamondsuit$

Let then $K \subseteq A^* \cdot \$ \cdot A^+$ be a rational language saturated by $\overset{UP}{\equiv}$. Let $\mathcal{A} = (Q, I, F, E)$ a deterministic automaton which recognizes K. We denote by δ the transition function of the automaton $\mathcal{A}$ and let q_0 be its initial state. We set $Q_d = \{q \in Q \mid \exists u \cdot \$ \cdot v \in K, q = \delta(q_0, u)\}$. For each state $q \in Q_d$ we denote by M_q the language of words u such that $\delta(q_0, u) = q$, and we denote by N_q the language of words v such that $\delta(q, v)$ if a final state. M_q and N_q are rational languages and $K = \cup_{q \in Q_d} M_q \cdot \$ \cdot N_q$ because K is a subset of $A^* \cdot \$ \cdot A^+$.

The language N_q is recognized by the automaton $\mathcal{A}_q = (Q, \{\delta(q, \$)\}, F, E)$, and for each final state q_f, we let the rational language N_{q,q_f} be the set of words v such that $\delta(q, v) = q$ and $\delta(q, \$ \cdot v) = q_f = \delta(q_f, v)$. This language is composed of words v of N_q that loop on both q and q_f, the final state of the calculus of $\mathcal{A}_q$ on v. Finally, we define the ω-rational language L by

$$L = \bigcup_{(q,q_f) \in Q_d \times F} M_q \cdot N^{\omega}_{q,q_f} \tag{2}$$

The languages M_q and N_{q,q_f} satisfy the hypothesis of Lemma 5 , i.e. $N^+_{q,q_f} = N_{q,q_f}$ and $M_q \cdot N^*_{q,q_f} = M_q$. Each ultimately periodic word α which is an element of $M_q \cdot N^{\omega}_{q,q_f}$ is equal to $u \cdot v^{\omega}$ with $u \in M_q$ and $v \in N_{q,q_f}$. Then, $u \cdot \$ \cdot v \in K$ and we deduce from the saturation of K by $\overset{UP}{\equiv}$ that all words $u \cdot \$ \cdot v$ such that $\alpha = u \cdot v^{\omega}$ are elements of K. We have thus shown the inclusion $L_\$ \subseteq K$.

Conversely, let $u \cdot \$ \cdot v$ be a word of K. For each integer k, words $u \cdot \$ \cdot v$ and $u \cdot v^k \cdot \$ \cdot v$ represent the same ultimately periodic word. K is saturated by $\overset{UP}{\equiv}$ thus, $u \cdot v^k \cdot \$ \cdot v \in K$ and $\delta(q_0, u \cdot v^k) \in Q_d$. Let the sequence of states $p_k \in Q_d$ be defined by $p_k = \delta(q_0, u \cdot v^k)$. Q_d is finite, so we can find two integers r and m such that $m \geq 1$, $p_r = p_{r+m}$ and for each integer $k \leq r + m$, $p_k \notin \{p_0, \ldots, p_{k-1}\}$. We may show by a simple induction that $p_{k+m} = p_k$ for each integer $k \geq r$. We set $r = sm + r'$, with $0 \leq r' < m$, and $k_1 = r + m - r' = (s+1)m$. Then, we get $p_{2k_1} = p_{k_1+(s+1)m} = p_{k_1}$, because $k_1 > r$, and then $q_0 \underset{\mathcal{A}}{\xrightarrow{u \cdot v^{k_1}}} p_{k_1} \underset{\mathcal{A}}{\xrightarrow{v^{k_1}}} p_{k_1}$. We set $q = p_{k_1}$. With a similar argument on the sequence of final states p'_k defined by $p'_k = \delta(q, \$ \cdot (v^{k_1})^k)$, we show that there exists an integer k_2 such that $p \underset{\mathcal{A}}{\xrightarrow{\$ \cdot v^{k_1 k_2}}} p'_{k_2} \underset{\mathcal{A}}{\xrightarrow{v^{k_1 k_2}}} p'_{k_2}$. We set $q_f = p'_{k_2}$. We have thus showed that $u \cdot v^{\omega} \in M_q \cdot N^{\omega}_{q,q_f}$, because $u \cdot v^{\omega} = u \cdot v^{k_1} \cdot (v^{k_1 k_2})^{\omega}$ and the words $u \cdot v^{k_1}$ and $v^{k_1 k_2}$ are in M_q and in N_{q,q_f}, respectively . The infinite word $u \cdot v^{\omega}$ is in L, and this proves the set inclusion $K \subseteq L_\$$. Finaly, we have showed the following proposition.

Proposition 6 *Let $K \subseteq A^* \cdot \$ \cdot A^+$ a rational language. Then, there exists a rational ω-language L such that $K = L_\$$ if and only if K is saturated by the equivalence $\overset{UP}{\equiv}$.* $\diamond$

We can build directly from $\mathcal{A}$ an automaton recognizing the ω-language L. The set Q_d can be effectively computed. For each state $q \in Q_d$, the language M_q is recognized by the automaton $(Q, I, \{q\}, E)$, which have m states. For each final state q_f, the language N_{q,q_f} is the intersection of the tree languages $L(Q, \{q\}, \{q\}, E)$, $L(Q, \{\delta(q, \$)\}, \{q_f\}, E)$ and $L(Q, \{q_f\}, \{q_f\}, E)$, and this language is recognized by an automaton with m^3 states. Each ω-language $M_q \cdot N_{q,q_f}^{\omega}$ is thus recognized by an automaton with $m + m^3$ states. There are at most m^2 pairs $(q, q_f) \in Q_d \times F$, and then the ω-language L is recognized by an automaton which has at most $m^3 + m^5$ states.

7 Remarks

The set $\hat{K} = \{u \cdot v^{\omega} \mid u \cdot \$ \cdot v \in K\}$ of ultimately periodic words corresponding to a rational set K of finite words in $A^* \cdot \$ \cdot A^+$ needs not be equal to $UP(M)$ for any rational language $M \in \mathcal{R}at(A^{\omega})$. In fact, there exists $M \in \mathcal{R}at(A^{\omega})$ such that $\hat{K} = UP(M)$ if and only if the smallest language containing K saturated by $\overset{UP}{\equiv}$ is rational, and this is not always the case.

For example, $K = \$ \cdot A^+$ is a rational set of finite words include in $A^* \cdot \$ \cdot A^+$. $\hat{K}$ is the set of periodic words on the alphabet A and $(\hat{K})_\$$ is not a rational set if A has more than one letter. In fact, if a and b are distinct letters of A, $\$ \cdot a \cdot b^n \in K$ for each $n \in \mathrm{N}$ and then $a \cdot b^n \cdot \$ \cdot a \cdot b^n \in (\hat{K})_\$$ for each integer n. But for each integer n', $n' < n$, $a \cdot b^{n'} \cdot \$ \cdot a \cdot b^n \notin (\hat{K})_\$$ because the word $a \cdot b^{n'} \cdot (a \cdot b^n)^{\omega}$ is not periodic. The language $(\hat{K})_\$$ may not be rational because it does not even satisfy pumping lemma conclusions.

8 Conclusion

We have solved in principle the problem of building an effective one-to-one correspondance between Büchi automata and dfa's recognizing the languages $UP(M)_\$$. This raises two immediate natural questions. How can we decide efficiently that a rational language $K \subseteq A^* \cdot \$ \cdot A^+$ is saturated by $\overset{UP}{\equiv}$? How can we decide that $(\hat{K})_\$$ is rational ? More generally, the question is raised to derive from canonical forms of the dfa's recognizing the $UP(M)_\$$ canonical forms for the Büchi automata recognizing the M's and hopefully efficient practical algorithms for the manipulation of Büchi automata.

9 Bibliography

[1] Arnold, A., "A Syntactic Congruence for Rational ω-Languages," *Theoretical Computer Science,* **39**, (1985), 333-335.

[2] Calbrix, H., Nivat, M., Podelski, A., "Une méthode de décision de la logique monadique du second ordre d'une fonction successeur", submited to *Comptes Rendus de l'Académie des Sciences, Série I.*

[3] Calbrix, H., Nivat, M., Podelski, A., "Sur les mots ultimement périodiques des langages rationnels de mots infinis", submited to *Comptes Rendus de l'Académie des Sciences, Série I.*

[4] Eilenberg, S., "Automata, Languages and Machines," Vol. A, *Academic Press,* (1974).

[5] Safra, S., "On the complexity of ω-automata," in: *Proc. 29th Ann. IEEE Symp. on Foundations of Computer Science,* (1988), 319-327.

[6] Sistla, A.P., M.Y. Vardi and P. Volper, "The Complementation Problem for Büchi automata with Application to Temporal Logic," *Theoret. Comput. Sci.,* **49**, (1987), 217-237.

[7] Thomas, W., "Automata on Infinite Objects," in *Handbook of Theoretical Computer Science, J. van Leeuven ed., Elsevier,* (1990), 133-191.