An Algebraic Characterization of Finite Automata on Infinite Words

Zaynab Ghazi

April 28, 2022

Submitted to the Faculty of Bryn Mawr College

in partial fulfillment of the requirements for

Bacherlor's Degree in Computer Science.

Supervised by Prof. Steven Lindell.

Abstract

This paper presents a comprehensive summary of the study of finite automata operating on infinite inputs. We attempt to provide both mathematicians and computer scientists a more intelligible introduction to the topic of algebraic automata theory, mainly through presenting crucial results and findings in the topic, as well as extended proofs to what has only been sketched in existing papers. Our study will focus on the recognition of infinite words by Büchi automata, as well as $\omega-$semigroup and Wilke algebra morphisms, which constitute the algebraic equivalent of the automata. We also present crucial results in Ramsey theory that facilitate the transitions between the finite and the infinite, in order to generalize known results for finite words to infinite words, such as Kleene's theorem, and prove the equivalence of the aforementioned modes of recognition.

Acknowledgements

I would like to express my sincere gratitude to my supervisor Prof. Steven Lindell for introducing me to the topic of this thesis, and for his dedicated guidance and feedback which were crucial to writing this document. Our meetings and discussions propelled me to think outside the box and observe under a new light the intricate relations between computer science and mathematics, as well as the fascinating ways in which complex mathematical theorems can be directly derived from the most simplistic of axioms. I am also grateful to Prof. Aline Normoyle for her insightful weekly feedback during senior conference.

Finally, I would also like to thank my family and friends who still wonder why I did not build a mobile app instead.

Table of Contents

Chapter 1

Introduction

1.1 Background

The theory of computation constitutes one of the core disciplines of computer science. Most computer scientists are exposed at the undergraduate level to the mechanisms of finite automata and the basics of formal languages. However, this exposure is often restricted to the behavior of automata on finite words. This limitation is due in large part to the level of difficulty introduced by studying infinite objects as compared to finite ones. Nonetheless, most results known for finite words apply to infinite words. This has been proven by Büchi in the early 1960s. He also discovered a connection between sets of models of formulas in the monadic second order logic of infinite sequences and ω-regular languages [6]. The algebraic approach, which this paper attempts to present, was subsequently introduced by Schützenberger who established an equivalence between star-free languages and finite aperiodic syntactic monoids in the 1990s, rendering the theory of finite automata closely tied to that of finite monoids [8].

While monoid theory was very successful in the recognition of finite words, the transition to infinite words required a different algebraic framework since every product in the monoid structure of $A^{\mathbb{N}}$ (*i.e.* the language of infinite words) is equal to its first factor. Furthermore, the astounding similarity between A^* (*i.e.* the language of finite words) and $A^{\mathbb{N}}$ implied that the study of infinite words is inseparable from that of their finite counterparts, which pushed computer scientists to resort to multi-sorted algebras where operations can be defined on both spaces simultaneously [10]. For instance, $\omega-$semigroups which are discussed later in the paper support a *mixed product* of the form **finite** $\times$ **infinite** $\rightarrow$**infinite**. They also provide a natural transition from finite words to infinite ones through defining a product of infinite arity over finite words. Similarly, Wilke algebras, which are also an algebraic mode of recognition of infinite words, support a similar transition by applying the unary ω operation to finite words.

1.2 Motivation and Purpose

Algebraic automata theory not only bridges two fields of inquiry of the utmost importance, notably **set theory** and **the theory of computation**; but it also has applications that extend to practical aspects of computer science such as software verification, word processing and network and compiler design. In fact, the operation of finite automata on infinite words is one of the most successful ways to verify the performance and correctness of a myriad of software systems, especially reactive ones such as operating systems and network servers, which are also concurrent systems whose implementation is error-prone due to concurrency rendering bugs difficult to reproduce. This procedure also extends to hardware systems such as airplanes and industrial machines wherein errors can be fatal, and a rigorous framework of verification is crucial. The process of representing these systems as automata and exploring their state space to check for violations of the desired system specifications is known as **model checking**, and is employed extensively in the software industry [4]. For instance, ***JavaPathFinder*** is an open-source model checker that allowed NASA to detect multiple subtle bugs in complex Java code, and ***Cmc*** is a model checker for C-programs employed in the verification of file systems and networks [5]. It is worth mentioning that algorithms on automata can be of extremely high complexity, thereby cultivating more interest in their algebraic characterization which often leads to more efficient solutions [10].

On these grounds, this paper will present the algebraic characterization of the behavior of finite automata on infinite words as follows:

- In Section 2.1, we introduce the standard notation and definitions relating to formal languages and automata, then define the recognition of infinite words by finite automata.
- In Section 2.2, we introduce algebraic structures recognizing infinite words, such as semigroups, $\omega-$semigroups and Wilke algebras.
- In Section 2.3, we present the morphisms enabling algebraic recognition.
- In Section 2.4, we present a method to convert a Büchi automaton to an $\omega-$semigroup.
- In Chapter 3, we prove the equivalence between all the modes of recognition of infinite words presented in this paper.
- In Chapter 4, we conclude our paper with the major points discussed therein.

Chapter 2

Preliminaries

In this chapter, we shall define the notions from automata theory and algebra necessary to our study. We shall also present and prove some key results regarding the equivalence of these notions.

2.1 Formal languages and automata

Automata constitute an intuitive and fundamental representation of a machine, or the more abstract notion of a computation. They are equivalent to a sequential input processing algorithm where the input is a sequence of symbols belonging to a defined alphabet denoted $\sum$, and whose finality is the recognition of specific strings/words belonging to a defined language; hence their extensive use in pattern matching and text analysis. Each symbol of the alphabet transitions the automaton from a state to another. This paper will focus on automata with a finite number of states, also known as finite automata. We define a finite automaton as follows:

Definition 2.1. A **finite automaton** M is a 5-tuple $(Q, \sum, \delta, I, F)$ where

1. Q is the finite set of possible states of M,
2. $\sum$ is a the alphabet of M,
3. $\delta : Q \times \sum \rightarrow \mathcal{P}(Q)$ is the transition function of M,
4. $I \subseteq Q$ is the set of initial states of M, and
5. $F \subseteq Q$ is the set of final/accept states of M.

Since we are concerned with the behavior of finite automata on the set of infinite words (denoted $\sum^{\omega}$), we shall focus on a specific type of automaton called the **Büchi automaton**. We define recognition on such an automaton as follows:

Definition 2.2. [1] $B = (Q, \sum, \delta, I, F)$ be a Büchi automaton and
$w = w_1 w_2 w_3 ...$ be an infinite word on $\sum$.
A **run** of B on w is a sequence of the form $q_1 q_2 q_3 ...$ such that $q_1 \in I$, and $q_{i+1} \in \delta(q_i, w_i)$, $\forall i \in \mathbb{N}$.
We denote all the possible runs of B on w as $\Delta(w)$.
We say B **accepts** w if there is run in $\Delta(w)$ that visits an accepting state infinitely often.
Formally, Let $p \in \Delta(w)$. The set of infinitely repeated states in p is denoted $Inf(p)$.
Then, B **accepts** w if $\exists p \in \Delta(w)\, s.t.\, Inf(p) \cap F \neq \emptyset$.
If B accepts w, we write $w \in L^{\omega}(B)$, where $L^{\omega}(B)$ is the language recognized by the automaton B.

Example. We present an example of a Büchi automaton used in model checking. This automaton (B_1) depicts the communication between a process P and a server S, such that P can only send a new request after its last one has been granted by S. An example of a specification we would be interested in verifying is that every request will be granted eventually.

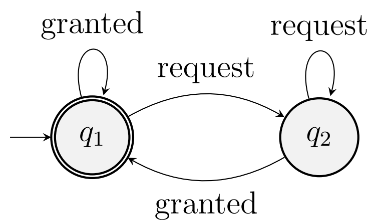

Figure 2.1: Büchi automaton depicting a simple process-server communication [3].

In the formal definition, $B = (\{q_1, q_2\}, \{\text{request,granted}\}, \delta, \{q_1\}, \{q_1\})$, where the transition function δ is given by the table:

	request	granted
q_1	q_2	q_1
q_2	q_2	q_1.

From the definition of acceptance above, we know B will only accept an infinite sequence of input over {request,granted} if it visits an accepting state infinitely often. The only accepting state is q_1 since $F = \{q_1\}$. From the transition function δ, q_1 can only be accessed after at least one request has been granted. Therefore, the runs accepted by B correspond to infinite sequences of a finite number of requests followed by the granting of at least one of them. This guarantees that every request will be eventually followed by a granted event.

We write $L^\omega(B) = (\text{request}^*\text{granted})^\omega$, where $*$ is Kleene's star and $()^\omega$ is the infinitization operator such that $\sum^*$ and $\sum^\omega$ denote the set of finite words and the set of infinite words, respectively, on the alphabet $\sum$. Their union is denoted $\sum^\infty$.

Consequently, the communication between the process and the network only satisfies the specification above if it can be represented as an infinite word that belongs to

$L^{\omega}(B)$.

We say (request*granted)$^{\omega}$ is a **recognizable language** because there exists a Büchi automaton that recognizes it. More generally, we say a set I of infinite words is recognizable if there exists B a Büchi automaton such that $L^{\omega}(B) = I$. I is also called an $\omega-$rational language by extension of Kleene's theorem to infinite words [1]. We define rationality and its analogue for infinite words, notably $\omega-$rationality as follows:

Definition 2.3. Given an alphabet $\sum$, the collection of rational/regular languages is defined by:

1. The empty set $\emptyset$ is rational.
2. Each symbol $\sigma \in \sum$ is rational.
3. If X and Y are rational, then the union $X \cup Y$, the concatenation $X \cdot Y$ and the the Kleene star application $X \rightarrow X^* = \{x_1...x_n | n \geq 0 \text{ and } x_1, ..., x_n \in X\}$ are rational.

Note that this class of languages corresponds exactly to the one recognized by deterministic finite automata (DFA's), with a DFA being a special case of a finite automaton where δ is defined $\delta : Q \times \sum \rightarrow Q$ and $|I| = 1$. Nonetheless, every DFA has an equivalent finite automaton (NFA).

This equivalence between the two classes of languages results from Kleene's theorem as stated below.

Theorem 2.1. ***Kleene's Theorem.***

Let $\sum$ be an alphabet and $X \subseteq \sum^*$.

Then X is rational if and only if it is recognized by a DFA.

This notion of rationality can also be extended to infinite words as described below.

Definition 2.4. [7] Let $\sum$ be an alphabet.
L is an $\omega-$rational subset of $\sum^{\omega}$ if and only if it is a finite union of sets of the form XY^{ω} where X and Y are rational subsets of $\sum^*$.

Now, let us look at some properties of languages recognized by Büchi automata, as they exhibit the same closure patterns as regular languages.

Theorem 2.2. ***Closure under union.***
Let $\sum$ be an alphabet and $X, Y \subseteq \sum^{\omega}$.
Then X and Y are Büchi recognizable $\implies X \cup Y$ is Büchi recognizable.

Proof. We have X is Büchi recognizable $\implies \exists B_X = (Q_X, \sum, \delta_X, I_X, F_X)$ a Büchi automaton such that $L^{\omega}(B_X) = X$.
Similarly, Y is Büchi recognizable $\implies \exists B_Y = (Q_Y, \sum, \delta_Y, I_Y, F_Y)$ a Büchi automaton such that $L^{\omega}(B_Y) = Y$.
We shall show $\exists B$ a Büchi automaton such that $L^{\omega}(B) = X \cup Y$.
Suppose without loss of generality that $Q_X \cap Q_Y = \emptyset$.
We let $B = (Q_X \cup Q_Y \cup \{q_I\}, \sum, \delta, \{q_I\}, F_X \cup F_Y)$ such that δ is defined as follows:

$$\delta(q,a) = \begin{cases} \{\delta_X(q',a) |\, q' \in I_X\} \cup \{\delta_Y(q',a) |\, q' \in I_Y\} & \text{if } q = q_I, \\ \delta_X(q,a) & \text{if } q \in Q_X, \\ \delta_Y(q,a) & \text{if } q \in Q_Y. \end{cases}$$

Now, let us show $L^{\omega}(B) = X \cup Y$.

1. Let us prove $X \cup Y \subseteq L^{\omega}(B)$.

Recall a successful run is one that visits a final state infinitely often.

$u \in X \cup Y \implies B_X$ accepts u or B_Y accepts u.

$\implies \exists$ a successful run of B_X or B_Y on u.

$\implies \exists$ a successful run $r = \{q_i\}_{i \geq 0}$ s.t. $q_0 \in I_A \wedge q_{i+1} \in \delta_A(q_i, u_i)$, $\forall i \in \mathbb{N} \cup \{0\}$ and $A \in \{X, Y\}$.

$\implies \exists$ a successful run r' of B on u s.t. $r' = q_I r''$ where $r'' = \{q_i\}_{i \geq 1}$.

$\implies u \in L^\omega(B)$.

Therefore, $X \cup Y \subseteq L^\omega(B)$.

2. Let us prove $L^\omega(B) \subseteq X \cup Y$.

$u \in L^\omega(B) \implies \exists$ a successful run $r = \{q_i\}_{i \geq 0}$ s.t. $q_0 = q_I \wedge q_{i+1} \in \delta(q_i, u_i)$, $\forall i \in \mathbb{N} \cup \{0\}$.

$\implies \exists$ a successful run $r = \{q_i\}_{i \geq 0}$ s.t. $q_0 = q_I \wedge q_1 \in Q_A$ $\wedge\, q_{i+1} \in \delta_A(q_i, u_i)$, $\forall i \in \mathbb{N}$ and $A \in \{X, Y\}$.

$\implies \exists$ a successful run $r' = q_A\{q_i\}_{i \geq 1}$ s.t. $q_A \in I_A \wedge q_{i+1} \in \delta_A(q_i, u_i)$, $\forall i \in \mathbb{N}$ and $A \in \{X, Y\}$.

$\implies \exists$ a successful run of B_X or B_Y on u.

$\implies u \in X \cup Y$.

Therefore, $L^\omega(B) \subseteq X \cup Y$.

From 1 and 2, we conclude $L^\omega(B) = X \cup Y$.

Therefore, Büchi recognizable languages are closed under union. $\square$

The following theorem establishes a key relation between rational languages and $\omega-$rational languages.

Theorem 2.3. *Let $\sum$ be an alphabet and $X \subseteq \sum^\omega$.*

If X is a rational language, then X^ω is Büchi recognizable.

Proof. Let $\sum$ be an alphabet, and $X \subseteq \sum^*$ be a rational language.

By Kleene's theorem, there exists a DFA $M = (Q, \sum, \delta, \{q_0\}, F)$ such that $L(M) = X$. For simplicity, we suppose there is no incoming arrow to q_0. In order to prove X^ω is Büchi recognizable, we shall show there exists a Büchi automaton B such that $L^\omega(B) = X^\omega$.

We propose constructing B as follows: $B = (Q, \sum, \delta', \{q_0\}, \{q_0\})$ where δ' is defined as follows:

$$\delta'(q,a) = \begin{cases} \delta(q,a) & \text{if } \delta(q,a) \cap F = \emptyset, \\ \\ \{q_0\} \cup \delta(q,a) & \text{otherwise.} \end{cases}$$

We shall prove $L^\omega(B) = X^\omega$.

1. Let us first prove $X^\omega \subseteq L^\omega(B)$.

   Let $u \in X^\omega$ be arbitrary.

   $u \in X^\omega \implies u = u_1u_2 ... \; s.t. \; u_i \in X \;, \forall i \in \mathbb{N}.$

   $\implies u = u_1u_2 ... \; s.t.$ there exists an accepting path in M labeled $u_i, \forall i \in \mathbb{N}$.

   $\implies u = u_1u_2 ... \; s.t.$ there exists a computation on M $q_{i_0}q_{i_1}...q_{i_{|u_i|}}, \forall i \in \mathbb{N} \; s.t.$

   $1.q_{i_0} = q_0,$

   $2.\delta(q_{i_j}, u_{i_j}) = q_{i_{j+1}} \; \forall j \in [\![0, |u_i| - 1]\!],$

   $3.q_{i_{|u_i|}} \in F.$

   $\implies$ A run of B on u is the sequence $p = q_{0_0}q_{0_1}...q_{0_{|u_i|-1}}q_0q_{1_0}q_{1_1}...q_{1_{|u_1|-1}}q_0... \; s.t.$

   1 and 2 hold,

i.e.

$$\begin{aligned} u \in X^\omega &\implies \exists p \in \Delta(u)\ s.t.\ Inf(p) \cap \{q_0\} \neq \emptyset. \\ &\implies u \in L^\omega(B). \end{aligned}$$

Therefore, we conclude $X^\omega \subseteq L^\omega(B)$.

2. Now, let us show $L^\omega(B) \subseteq X^\omega$.

$$\begin{aligned} u \in L^\omega(B) &\implies \exists p \in \Delta(u)\ s.t.\ Inf(p) \cap \{q_0\} \neq \emptyset. \\ &\implies \text{there exists a sequence } \{s_i\}_{i\in\mathbb{N}} \text{ and } p = p_1p_2... \in \Delta(u)\ s.t.\ p_{s_i} = q_0\ \forall i \in \mathbb{N}. \\ &\implies u = u_1u_2...\ s.t.\ u_i \text{ is the label of the path } p_{s_i}p_{s_{i+1}}\ \forall i \in \mathbb{N}. \\ &\implies u = u_1u_2...\ s.t. \text{ M accepts } u_i. \\ &\qquad (\text{since } p_{s_{i+1}} = q_0 \implies \delta(p_{s_{i+1}-1}, u_{i_{|u_i|-1}}) \cap F \neq \emptyset \text{ for all } i \geq 1) \\ &\implies u \in X^\omega. \end{aligned}$$

Therefore, we conclude $L^\omega(B) \subseteq X^\omega$.

From 1 and 2, we get $L^\omega(B) = X^\omega$ as desired.

□

2.2 $\omega-$semigroups and Wilke Algebras

Algebraic recognizability offers many advantages over automata, not only does it provide elegant and rigorous solutions to what can only be solved inefficiently and with unnecessary complexity by automata, but it also provides a much simpler way of classification of languages that facilitates the decision of membership in significant linguistic classes [10]. In this section, we shall define the primary algebraic notions used to recognize infinite words.

While the algebraic recognition of finite words relies on algebras such as semigroups and monoids, the study of infinite words requires multi-sorted algebras as some operations are defined on more than one algebra. This is also due to the fact that the study of infinite words is best not separated from that of finite words, especially that the former can be generated from the latter. In this section, we shall define these multi-sorted algebras, notably $\omega-$semigroups and Wilke algebras.

Definition 2.5. [1] An $\boldsymbol{\omega-}$**semigroup** is a two-sorted algebra $S = (S_+, S_\omega)$ equipped with the following operations:

1. an associative product on S_+ denoted $\cdot$,
2. a mapping $* : S_+ \times S_\omega \to S_\omega$ called *mixed product*,
3. a surjective mapping $\pi : S_+^\omega \to S_\omega$ called *infinite product.*

The $\omega-$semigroup can also be denoted $S = (S_+, S_\omega, \cdot, *, \pi)$. Furthermore, the infinite product is consistent with the two other operation, such that for every increasing sequence $(k_n)_{n>0}$ and for every sequence $(s_n)_{n\geq 0}$ of elements in S_+,

$$\pi(s_0 s_1 \ldots s_{k_1-1}, s_{k1} s_{k_1+1} \ldots s_{k_2-1}, \ldots) = \pi(s_0, s_1, s_2, \ldots), \tag{2.1}$$

and for every element $s \in S_+$, we have

$$s\pi(s_0, s_1, s_2, \ldots) = \pi(s, s_0, s_1, s_2, \ldots). \tag{2.2}$$

Note that (2.1) and (2.2) are but an extension of associativity that allows writing the product $\pi(s_0, s_1, s_2, \ldots)$ unambiguously as $s_0 s_1 s_2 \ldots$ since all possible factorizations of the product are equivalent.

Example. (The Free ω-semigroup over $\sum$)

Defining the product operations as concatenation and given an alphabet $\sum$, we get the $\omega-$semigroup $\sum^\infty = (\sum^*, \sum^\omega, \cdot, *, \pi)$ where

- $\cdot$ represents the concatenation of two finite words over $\sum^*$,
- $*$ represents concatenating a finite word and an infinite word, naturally resulting in an infinite word over $\sum^\omega$,
- π represents the concatenation of an infinite number of finite words, thus generating an infinite word.

Now, we shall define the second multi-sorted algebra used in infinite word recognition. It was defined by Wilke in an attempt to provide a finite signature for the infinite product π as defined in an $\omega-$semigroup.

Definition 2.6. [1] A **Wilke algebra** is a two-sorted algebra $S = (S_+, S_\omega, \cdot, *, ()^\omega)$ where

- $\cdot$ is an associative multiplication on S_+,
- $* : S_+ \times S_\omega \to S_\omega$, called the *mixed product*, satisfies $s(t * u) = (st) * u$ for all $s, t \in S_+$ and $u \in S_\omega$,
- $()^\omega : S_+ \to S_\omega$ maps each element $s \in S_+$ to $s^\omega \in S_\omega$ and satisfies the following properties for all $s, t \in S_+$:

  1. $s(ts)^\omega = (st)^\omega$,
  2. $(s^n)^\omega = s^\omega$ for all $n > 0$.

Note that the notion of an $\omega-$semigroup can be made equivalent to that of a Wilke algebra by setting the infinite product $\pi(s, s, ...) = sss... = s^\omega$.
In the rest of this paper, we shall omit the symbol for the product operations denoted $*$ and $\cdot$ since they are jointly associative.

2.3 Algebraic recognition by morphisms

Before the algebraic recognition of infinite words was defined on the notion of $\omega-$semigroups and Wilke algebras, it was studied analogously to the study of the recognition of finite words, which is achieved through morphisms onto finite semigroups and monoids. A semigroup is a pair $(S, \cdot)$ where $\cdot$ is an associative binary operation on S. If S contains the identity of $\cdot$, it is called a *monoid*. This early notion of recognizability is called **strong recognition**. In order to define it, we shall define relevant algebraic notions.

Definition 2.7. Let $(S, \cdot)$ be a finite semigroup.
An element e is said to be *idempotent* if $e^2 = e$.
A pair (s, e) is said to be a *linked pair* if e is idempotent and $se = s$.

The notion of linked pairs is crucial to the recognizability of infinite words as a result of Ramsey's theorem, which provides a finite representation of infinite words in terms of these pairs called a *Ramseyan factorization*. Let us first introduce Ramsey's theorem.

Theorem 2.4. *Let $k \in \mathbb{N}$.*
Given any map $c : \mathbb{N}^{(2)} \to [k]$, there exists $X \subseteq \mathbb{N}$ such that X is infinite and $c|_{X^{(2)}}$ is a constant function, where $c|_{X^{(2)}}$ is the restriction of c to $X^{(2)}$.

Note that c is usually known as a **coloring function** and the notation $[k]$ is equivalent to $[\![1, k]\!]$.

Proof. Let $k \in \mathbb{N}$ and let $c : \mathbb{N}^{(2)} \to [k]$ be an arbitrary coloring function.

We shall construct $X \subseteq \mathbb{N}$ such that X is infinite and $c|_{X^{(2)}}$ is a constant function, that is to say all the pairs over X have the same color.

Let $\alpha_1 \in \mathbb{N}$ be arbitrary. By the Pigeonhole Principle (P.P.), there exists an infinite set $\beta_1 \subseteq \mathbb{N} \setminus \{\alpha_1\}$ such that all the pairs $\{(\alpha_1, b_1)|b_1 \in \beta_1\}$ are of the same color, *i.e.* the set $\{(\alpha_1, b_1)|b_1 \in \beta_1\}$ is monochromatic.

Similarly, we pick $\alpha_2 \in \beta_1$ randomly. Then, there exists an infinite set $\beta_2 \subseteq \beta_1 \setminus \{\alpha_2\}$ such that the set $\{(\alpha_2, b_2)|b_2 \in \beta_2\}$ is monochromatic.

We proceed inductively until we obtain a sequence $\{\alpha_1, \alpha_2, ...\}$ of elements in $\mathbb{N}$ and a sequence $c_1, c_2, ...$ of colours such that $c(\alpha_i, \alpha_j) = c_i$ for all $i < j$.

Again, by the P.P., we know there are finitely many colors in $[k]$ and therefore, there must exist a sequence $c_{\delta_1}, c_{\delta_2}, ...$ such that $c_{\delta_1} = c_{\delta_2} = ... = c \in [k]$.

Let $X = \{\alpha_{\delta_i}|i \in \mathbb{N}\}$. Then $X^{(2)}$ is monochromatic since $c|_{X^{(2)}} = c$, *i.e.* $c|_{X^{(2)}}$ is constant as desired. $\square$

Now, we shall present and prove the application of Ramsey's theorem as stated above to infinite words.

Theorem 2.5. *[1]*

Let $\varphi : \sum^+ \to S$ be a morphism from $\sum^+$ into a finite semigroup S.

For all infinite words $u \in \sum^\omega$, there exist a linked pair (s, e) and a factorization $u = u_0u_1u_2...$ as a product of words in $\sum^+$ such that $\varphi(u_0) = s$ and $\varphi(u_i) = e$ for all $i > 0$.

Proof. This statement can be proved by applying Ramsey's theorem to the map $\varphi : \mathbb{N}^{(2)} \to S$. We can identify the substring $u[i, j] \in \sum^+$ with the subset $\{i, j\} \in \mathbb{N}^{(2)}$. By Ramsey's theorem, we conclude there exists an infinite set $X \subseteq \mathbb{N}$ such that $\varphi|_{X^{(2)}}$ is constant, $i.e. \forall i, j \in X \; s.t. \; i < j, \varphi(u[i, j]) = e \in S$. (*)

Then, let $\{\delta_i\}_{i\geq 0}$ be the sequence of elements of X sorted in ascending order.

We can then let $u_0 = u[0, \delta_1]$ and $u_i = u[\delta_i, \delta_{i+1}]$ for all $i > 0$.

Consequently, we will have $\varphi(u_i) = e$ for all $i > 0$.

Then, we let $s = \varphi(u[0, \delta_1]) = \varphi(u[0, \delta_0])\varphi(u[\delta_0, \delta_1]) = re.$ $(\diamond)$

Now, it suffices to prove (s, e) is a linked pair, *i.e.* we shall show $e^2 = e$ and $se = s$.

By (*), we know $e = \varphi(u[\delta_i, \delta_{i+1}]) = \varphi(u[\delta_{i+1}, \delta_{i+2}])$ for all $i \geq 0$

1. Then

$$\begin{aligned} e^2 &= ee \\ &= \varphi(u[\delta_i, \delta_{i+1}])\varphi(u[\delta_{i+1}, \delta_{i+2}]) \\ &= \varphi(u[\delta_i, \delta_{i+2}]) \\ &= e. && \text{(by *)} \end{aligned}$$

2. Similarly,

$$\begin{aligned} se &= \varphi(u[0, \delta_1])\varphi(u[\delta_1, \delta_2]) \\ &= \varphi(u[0, \delta_0])\varphi(u[\delta_0, \delta_1])\varphi(u[\delta_1, \delta_2]) \\ &= ree \\ &= re && \text{(since ee=e)} \\ &= s. && \text{(by } \diamond) \end{aligned}$$

Therefore, $e^2 = e$ and $se = s$ as desired. Hence, (s, e) is a linked pair.

We conclude that for all infinite words $u \in \sum^{\omega}$, there exist a linked pair (s, e) and a factorization $u = u_0u_1u_2...$ as a product of words in $\sum^{+}$ such that $\varphi(u_0) = s$ and $\varphi(u_i) = e$ for all $i > 0$. $\square$

Consequently, we call $\varphi-$simple a set of infinite words of the form $\varphi^{-1}(s)(\varphi^{-1}(e))^{\omega}$, and say that a subset of $\sum^{\omega}$ is **weakly recognized** if it is a finite union of $\varphi-$simple subsets [1]. Moreover, if the union is closed under conjugation of the linked pairs that generate it, we say that it is **strongly recognized** [2]. The latter mode of recognition significantly simplifies the process of complementation. More formally, we define strong recognition as follows:

Definition 2.8. [1] Let $\varphi : \sum^{+} \to S$ be a morphism from $\sum^{+}$ into a finite semigroup S.

Then φ strongly recognizes $X \subseteq \sum^{\omega}$ if for every linked pair (s, e) of S,

$$\varphi^{-1}(s)(\varphi^{-1}(e))^{\omega} \cap X = \emptyset \text{ or } \varphi^{-1}(s)(\varphi^{-1}(e))^{\omega} \subseteq X.$$

Now, we shall define the most recent notion of recognizability of infinite words, notably $\omega-$semigroup morphisms:

Definition 2.9. [1],[7] Given two $\omega-$semigroups $S = (S_+, S_\omega)$ and $T = (T_+, T_\omega)$, a morphism of $\omega-$semigroups is a pair $\varphi = (\varphi_+, \varphi_\omega)$ where $\varphi_+ : S_+ \to T_+$ is a semigroup morphism and the mapping $\varphi_\omega : S_\omega \to T_\omega$ satisfies

$$\varphi_\omega(s_0 s_1 s_2 ...) = \varphi_+(s_0)\varphi_+(s_1)\varphi_+(s_2)...$$

We use this definition in order to define recognizability of $\omega-$languages by $\omega-$semigroups as follows:

Definition 2.10. Let $\varphi : \sum^{\infty} \to S = (S_+, S_\omega)$ be a morphism of $\omega-$semigroups.

We say φ recognizes a subset X of $\sum^{\omega}$ if there exists a pair $P = (P_+, P_\omega)$ with $P_+ \subseteq S_+$ and $P_\omega \subseteq S_\omega$ such that $X = \varphi^{-1}(P)$,

i.e. $X_+ = X \cap \sum^{+} = \varphi_+^{-1}(P_+)$, and $X_\omega = X \cap \sum^{\omega} = \varphi_\omega^{-1}(P_\omega)$.

Example. Let $\sum = \{a, b, c\}$, and consider the $\omega-$semigroup $S = (\{0, 1, -1\}, \{0^\omega, 1^\omega\})$ equipped with the operations

$00 = 01 = 10 = (-1)0 = 0(-1) = 0$, $11 = (-1)(-1) = 1$, $(-1)1 = 1(-1) = -1$,

$11^\omega = 1^\omega, 00^\omega = 10^\omega = (-1)0^\omega = 01^\omega = 0(-1)^\omega = 0^\omega$ and

$1(-1)^\omega = (-1)(-1)^\omega = (-1)1^\omega = (-1)^\omega = 1^\omega$.

Let $\varphi : \sum^\infty \to S$ be the morphism of $\omega-$semigroups defined by $\varphi(a) = 0, \varphi(b) = 1$ and $\varphi(c) = -1$. Then

$\varphi^{-1}(0) = \sum^* a \sum^*$ is the set of finite words containing at least one a,

$\varphi^{-1}(1) = (b \cup (cb^*c))^+$ is the set of finite words that don't contain any a's and can contain any number of b's and an even number of c's,

$\varphi^{-1}(-1) = b^*c(b \cup cb^*c)^*$ is the set of finite words that don't contain any a's and contain an odd number of c's and any number of b's,

$\varphi^{-1}(0^\omega) = \sum^\omega \setminus (b^\omega \cup c^\omega)$ is the set of infinite words containing at least 1 a,

$\varphi^{-1}(1^\omega) = (\sum \setminus \{a\})^\omega$ is the set of infinite words containing no occurrence of a.

Therefore, the morphism φ recognizes each of these sets, as well as the union of any number of them.

2.4 Transition from Büchi automata to $\omega-$semigroups

In the previous sections, we presented different notions of recognizability of infinite words. In this section, we shall present a method to transition from a Büchi automaton to an $\omega-$semigroup prior to proving the equivalence of all these modes of recognition in the following chapter.

Let $B = (Q, \sum, \delta, I, F)$ be an arbitrary Büchi automaton. We start by constructing the *transition semigroup* of B denoted $\mathcal{S}(B)$.

To that end, we define φ to be the morphism that maps each symbol of the alphabet to its corresponding state mapping and each finite word to the paths for which it is a label, as follows:

$$\begin{aligned} \varphi : \Sigma^+ &\to \mathcal{R}(Q) \\ a &\mapsto \{(q, q') \in Q^2 | q' \in \delta(q, a)\} \\ u = u_1 ... u_n &\mapsto \{(q, q') \in Q^2 | q \xrightarrow{u} q'\} \end{aligned}$$

where $\mathcal{R}(Q)$ is the set of all binary relations defined on Q, and $q \xrightarrow{u} q'$ encodes the idea that there exists a path starting at state p and ending at state q that is labeled u.

Then, the transition semigroup of B is $\mathcal{S}(B) = \varphi(\Sigma^+)$.

Example. Let us apply this conversion method to the Büchi automaton from Figure 2.1. For simplicity, we let $\sum = \{r, g\}$ where $r = \textit{request}$ and $g = \textit{granted}$.

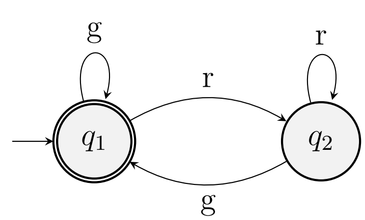

Figure 2.2: Büchi automaton depicting a simple process-server communication (simplified alphabet)[3].

The morphism φ is defined as follows:

$$\varphi(g) = \{(q_1, q_1), (q_2, q_1)\}, \varphi(r) = \{(q_1, q_2), (q_2, q_2)\}$$

$$\varphi(gr) = \{(q_1, q_2), (q_2, q_2)\}, \varphi(rg) = \{(q_1, q_1), (q_2, q_1)\}$$

$$\varphi(g^2) = \{(q_1, q_1), (q_2, q_1)\} \text{ and } \varphi(r^2) = \{(q_1, q_2), (q_2, q_2).\}$$

Therefore, the semigroup generated by these matrices contains two elements $\{g, r\}$ both idempotent, and is presented by the relations $rg = g$, $gr = r$, $r^2 = r$ and $g^2 = g$. Now, we shall extend the semigroup to a Wilke algebra recognizing $L^\omega(B)$. To that end, we shall define the following:

Definition 2.11. Let $(s, e), (s', e') \in \mathcal{L}$ be two arbitrary linked pairs of S.
(s, e) and (s', e') are said to be *conjugate* if there exist $x, y \in S^1$ such that $e = xy$, $e' = yx$, $s' = sx$ where S^1 is the monoid of S.
We write $(s, e) \sim (s', e')$.

Let $\bar{S} = (S_+, S_\omega, \cdot, *, ()^\omega)$ be the extension of the transition semigroup $\mathcal{S}$. Following the method described in [7] and [1], we set $S = \mathcal{S}(B)$ and S_ω to be the set of conjugacy classes of linked pairs of $\mathcal{S}(B)$.
Let $\mathcal{L}$ be the set of linked pairs. We define the mixed product $*$ for every conjugacy class $[s, e] \in S_\omega$ and every element $t \in S$ as follows:

$$t * [s, e] = t[s, e] = [ts, e].$$

Then, we define $()^\omega$ as follows:

$$()^\omega : t \to t^\omega = [t^\pi, t^\pi]$$

where π is the exponent of S, *i.e.* π is the smallest $n \in \mathbb{N}$ that renders s^n idempotent for all $s \in S$. This is a consequence of the following proposition.

Proposition 2.6. *[7] For each finite semigroup S, there exists an integer π, such that for each $s \in S$, s^π is idempotent.*

Now, let us prove this construction results in a well-defined Wilke algebra.

Proof. First, we shall prove that S_ω as constructed is well-defined, by proving that the conjugacy relation on S is an equivalence relation.
Let $\mathcal{R}$ be the conjugacy relation.

1. Let us show $\mathcal{R}$ is reflexive.

   Let $l \in \mathcal{L}$, we shall show $l\mathcal{R}l$.

   By definition, $l \in \mathcal{L} \implies \exists(s,e) \in S^2$ such that $se = s$ and $e^2 = e$.

   Let $x = y = e \in S$, then $e = ee = xy = yx$ and $se = sx = s$.

   By definition of conjugate pairs, we conclude $(s,e)\mathcal{R}(s,e)$,

   i.e. $l\mathcal{R}l$ as desired.

2. Let us show $\mathcal{R}$ is symmetric.

   Let $l, l' \in \mathcal{L}$, we shall show $l\mathcal{R}l' \implies l'\mathcal{R}l$.

   By definition, $l \in \mathcal{L} \implies \exists(s,e) \in S^2$ such that $se = s$ and $e^2 = e$.

   Similarly, $l' \in \mathcal{L} \implies \exists(s',e') \in S^2$ such that $s'e' = s'$ and $e'^2 = e'$.

   Suppose $l\mathcal{R}l'$. Then, there exist $x, y \in S^1$ such that $e = xy$, $e' = yx$, and $s' = sx$.

   Let $x' = y \in S$ and $y' = x \in S$. Then $e' = yx = x'y'$, $e = xy = y'x'$ and

$$
\begin{aligned}
s' = sx &\implies s' = sy' \\
&\implies s'x' = sy'x' \\
&\implies s'x' = s(y'x') \\
&\implies s'x' = se && \text{(since } e = y'x'\text{)} \\
&\implies s'x' = s. && \text{(since } se = s\text{)}
\end{aligned}
$$

   Therefore, $s = s'x'$, $e' = x'y'$ and $e' = y'x'$.

   By definition of conjugacy, $(s',e')\mathcal{R}(s,e)$,

   i.e. $l'\mathcal{R}l$ as desired.

3. Let us show $\mathcal{R}$ is transitive.

   Let $l, l', l'' \in \mathcal{L}$. We shall prove if $l\mathcal{R}l'$ and $l'\mathcal{R}l''$ then $l\mathcal{R}l''$.

Assume $l\mathcal{R}l'$ and $l'\mathcal{R}l''$.

By definition, $l \in \mathcal{L} \implies \exists(s,e) \in S^2$ such that $se = s$ and $e^2 = e$.

Similarly, $l' \in \mathcal{L} \implies \exists(s',e') \in S^2$ such that $s'e' = s'$ and $e'^2 = e'$ and

$l'' \in \mathcal{L} \implies \exists(s'',e'') \in S^2$ such that $s''e'' = s''$ and $e''^2 = e''$.

We have $l\mathcal{R}l'$. Then, there exist $x, y \in S^1$ such that $e = xy$, $e' = yx$, and $s' = sx$. (1)

Similarly, $l'\mathcal{R}l'' \implies$ there exist $x', y' \in S^1$ such that $e' = x'y'$, $e'' = y'x'$, and $s" = s'x'$. (2)

We shall show $l\mathcal{R}l''$,

i.e. show there exist $x'', y'' \in S^1$ such that $e = x''y''$, $e'' = y''x''$, and $s'' = sx''$.

From (1) and (2), we have $e' = e' \implies yx = x'y'$, (3)

we also have $s'' = s'x'$ and $s' = sx$. Therefore $s'' = sxx'$.

Let $x'' = xx' \in S^1$ since $x, x' \in S^1$. Then $s'' = sx''$.(4)

Let $y'' = y'y \in S^1$. Then

$$
\begin{aligned}
x''y'' &= xx'y'y \\
&= x(x'y')y \\
&= x(yx)y && \text{(by (3))} \\
&= (xy)(xy) \\
&= ee \\
&= e^2 = e, && \text{(since } e \text{ is idempotent)}
\end{aligned}
$$

i.e. $x''y'' = e$. (5)

Similarly, we have

$$\begin{aligned}
y''x'' &= y'yxx' \\
&= y'(yx)x' \\
&= y'(x'y')x' && \text{(by (3))} \\
&= (y'x')(y'x') \\
&= e''e'' \\
&= e''^2 = e'', && \text{(since } e'' \text{ is idempotent)}
\end{aligned}$$

i.e. $y''x'' = e''$. (6)

From (4), (5) and (6), we conclude there exist $x'', y'' \in S^1$ such that $s'' = sx''$, $e = x''y''$ and $e'' = y''x''$.

Therefore $(s, e)\mathcal{R}(s'', e'')$, *i.e.* $l\mathcal{R}l''$ as desired.

From 1,2 and 3, we conclude $\mathcal{R}$ is an equivalence relation, and therefore S_ω is well-defined.

Now, let us verify $\bar{S} = (S_+, S_\omega, \cdot, *, ()^\omega)$ is a Wilke algebra.

1. By construction, we know $(S_+, \cdot)$ is a semigroup since $(S_+, \cdot) = (S, \cdot)$.

2. Let us show $*$ is a mapping $S \times S_\omega \to S_\omega$ satisfying for all $s, t \in S$, and for every $u \in S_\omega$, $s * (t * u) = (s \cdot t) * u$.

   By construction, $* : S \times S_\omega \to S_\omega$.

   Let $s, t \in S$ and $u \in S_\omega$ be arbitrary.

   We have $u \in S_\omega \implies$ there exist $(s', e') \in \mathcal{L}$ such that $u = [s', e']$.

By construction,

$$t * u = [ts', e'] \implies s * (t * u) = [sts', e'] = [(st)s', e'] = (st) * [s', e'] = (s \cdot t) * u.$$

Therefore $s * (t * u) = (s \cdot t) * u$ for all $s, t \in S$ and every $u \in S_\omega$ as desired.

3. Let us now prove $()^\omega$ is a mapping $S \to S_\omega$ satisfying for all $s, t \in S$,

   $s(ts)^\omega = (st)^\omega$ and $(s^n)^\omega = s^\omega$ for all $n \geq 1$.

   Let $s, t \in S$ be arbitrary and $n \geq 1$.

   (a) We have $s(ts)^\omega = s * [(ts)^\pi, (ts)^\pi] = [s(ts)^\pi, (ts)^\pi]$.

   Then, to show $s(ts)^\omega = (st)^\omega$, it suffices to show $(s(ts)^\pi, (ts)^\pi)\mathcal{R}((st)^\pi, (st)^\pi)$.

   We know $s(ts)^\pi = (st)^\pi s$.

   Let $x = s \in S \subseteq S^1$. Then $s(ts)^\pi = (st)^\pi x$. (7)

   We also have $(st)^\pi = s(ts)^{\pi-1}t = x(ts)^{\pi-1}t$.

   Let $y = (ts)^{\pi-1}t$. Then $(st)^\pi = xy$. (8)

   Note $(ts)^\pi = (ts)^{\pi-1}(ts) = (ts)^{\pi-1}ts = yx$. (9)

   From (7), (8) and (9), we conclude there exist $x, y \in S^1$ such that

   $(st)^\pi = xy,\ (ts)^\pi = yx$ and $s(ts)^\pi = (st)^\pi x$.

   Therefore, $(s(ts)^\pi, (ts)^\pi)\mathcal{R}((st)^\pi, (st)^\pi)$,

   i.e. $s(ts)^\omega = (st)^\omega$.

   (b) Now, let us show $(s^n)^\omega = s^\omega$ for all $n \geq 1$.

   Let $n \geq 1$.

   We have $(s^n)^\omega = [(s^n)^\pi, (s^n)^\pi]$ and $s^\omega = [s^\pi, s^\pi]$.

   It suffices then to prove $(s^n)^\pi = s^\pi$ for all $n \geq 1$.

Using induction, we have the following:

If $n = 1$, then $(s^n)^\pi = s^\pi$ as desired.

Assume $(s^n)^\pi = s^\pi$ for some $n \geq 1$. Let us show $(s^{n+1})^\pi = s^\pi$.

We have $(s^{n+1})^\pi = (s^n)^\pi s^\pi = s^\pi s^\pi = s^{2\pi}$ by the inductive hypothesis.

Since π is the exponent of S, we get $s^{2\pi} = s^\pi$.

Therefore, $(s^{n+1})^\pi = s^\pi$. By the mathematical principle of induction, we conclude $(s^n)^\omega = s^\omega$ for all $n \geq 1$.

From $1, 2$ and 3, we conclude $\bar{S}$ is a Wilke algebra.

Example. Let us apply this construction on the semigroup corresponding to the automaton from Figure 2.2.

We know both g and r are idempotent, then the exponent of the semigroup is $\pi = 2$.

The possible linked pairs are $(g, g), (g, r), (r, r)$ and (r, g). Since $g = rg$ and $r = gr$, it is trivial that all these pairs are conjugate.

Therefore, $S_\omega = \{[g, r]\}$ and $r^\omega = g^\omega = [g, r]$.

Consequently, $\bar{S} = (\{g, r\}, \{[g, r]\})$. □

Chapter 3

Equivalence between Automata and Algebraic Recognition

The discussion of formal languages, automata and the algebraic recognition of infinite words leads us to the following theorem, which unifies all these notions.

Theorem 3.1. *Let $\sum$ be an alphabet and $X \subseteq \sum^{\omega}$. Then the following statements are equivalent:*

1. *X is $\omega-$rational.*

2. *There exists a Büchi automaton B such that $L^{\omega}(B) = X$.*

3. *X is strongly recognized by some morphism onto a finite semigroup.*

4. *X is recognized by some morphism onto a finite $\omega-$semigroup/Wilke algebra.*

Proof. In order to prove the theorem above, we shall prove the following:

$$1 \implies 2 \implies 3 \implies 4 \implies 1.$$

Let $\sum$ be an alphabet and $X \subseteq \sum^{\omega}$.

- **1** $\implies$ **2** : Suppose X is $\omega-$rational. Then, by definition 2.4, X is a finite union of subsets of the form ZY^ω where Z and Y are nonempty rational subsets of $\sum^*$. Let $X = \bigcup_{i=1}^n Z_i Y_i^\omega$, where $n \in \mathbb{N}$ and Z_i and Y_i are rational subsets of $\sum^*$ for all $i \in [\![1, n]\!]$.

  We shall then show $\exists B$ a Büchi automaton such that $L^\omega(B) = X$.

  To that end, let us prove that for all $i \in [\![1, n]\!]$, $\exists B_i$ a Büchi automaton such that $L^\omega(B_i) = Z_i Y_i^\omega$.

  Let $i \in [\![1, n]\!]$. By Theorem 2.2, we know Y_i^ω is Büchi recognizable. Therefore, $\exists B_{Y_i} = (Q_{Y_i}, \sum, \delta_{Y_i}, I_{Y_i}, F_{Y_i})$ a Büchi automaton s.t. $L^\omega(B_{Y_i}) = Y_i$.

  We know, by Kleene's theorem, Z_i is a rational language $\implies$ $\exists M_{Z_i}$ a finite automaton such that $L(M_{Z_i}) = Z_i$. For simplification, we take $M_{Z_i} = (Q_{Z_i}, \sum, \delta_{Z_i}, i_{Z_i}, f_{Z_i})$ to be a nondeterministic finite automaton with a unique initial state (i_{Z_i}) and a unique finite state (f_{Z_i}) such that there is no incoming arrow into i_{Z_i} and no outgoing arrow from f_{Z_i}.

  We construct a Büchi automaton $B_i = (Q_{Z_i} \cup Q_{Y_i}, \sum, \delta_i, i_{Z_i}, F_{Y_i})$ to recognize $Z_i Y_i^\omega$, where δ_i is defined as follows:

$$\delta_i(q, a) = \begin{cases} \delta_{Y_i}(q, a) & \text{if } q \in Q_{Y_i}, \\ \delta_{Z_i}(q, a) & \text{if } q \in Q_{Z_i} \text{ and } \delta_{Z_i}(q, a) \neq f_{Z_i}, \\ \delta_{Z_i}(q, a) \cup I_{Y_i} & \text{otherwise.} \end{cases}$$

  It it trivial to prove $L^\omega(B_i) = Z_i Y_i^\omega$.

  Therefore, $Z_i Y_i^\omega$ is Büchi recognizable $\forall i \in \mathbb{N}$. By Theorem 2.1, we conclude $\bigcup_{i=1}^n Z_i Y_i^\omega = X$ is Büchi recognizable, *i.e.* there exists a Büchi automaton B such that $L^\omega(B) = X$.

- **2** $\implies$ **3** : In section 2.4, we presented the transition semigroup of a Büchi automaton, however that only permits a weak recognition of the automaton's language [2]. We shall use a matrix-based approach presented in [2].

Suppose there exists a Büchi automaton $B = (Q, \sum, \delta, I, F)$ such that $L^\omega(B) = X$. We shall show X is strongly recognized by some morphism onto a finite semigroup, *i.e.* we shall construct a finite semigroup S and a morphism $\varphi : \sum^+ \to S$ such that for all linked pairs $(s, e) \in S^2$, $\varphi^{-1}(s)(\varphi^{-1}(e))^\omega$ has a trivial intersection with X.

For all $s \in S$, define $[s] := \varphi^{-1}(s)$.

Using notation form section 2.4, for $p, q \in Q$ and some finite word $u \in \sum^+$, we write $p \xrightarrow{u} q$ if there exists a sequence $q_0 u_0 ... q_{n-1} u_n q_n$ such that $p = q_0, q = q_n$ and $q_{i+1} \in \delta(q_i, u_i)$, $\forall i \in [\![0, n-1]\!]$.

Additionally, if $\exists i \in [\![0, n]\!]$ such that $q_i \in F$, we write $p \xrightarrow[F]{u} q$.

Accordingly, we construct φ as follows: $\varphi : \sum^+ \to K^{Q \times Q}$ where $K = \{0, 1, 2\}$,

and $\forall u \in \sum^+, \forall p, q \in Q, \varphi(u)_{pq} = \begin{cases} 1 & \text{if } p \xrightarrow{u} q, \\ 2 & \text{if } p \xrightarrow[F]{u} q, \\ 0 & \text{otherwise.} \end{cases}$

Let $S = \varphi(\sum^+)$.

For the purposes of this proof, a path $q_1 q_2 ... q_n$ on B is considered **valid** if $q_{i+1} \in \delta(q_i, u_i)\ \forall i \in [\![1, n-1]\!]$, and for all $i \in \mathbb{N}$ if the path is infinite.

1. Let us verify S is a semigroup.

   Note that the binary operation on S is defined as follows: $* : S \times S \to S$ such that $\varphi(x) * \varphi(y) = \varphi^*(xy)$ where $\varphi^*(xy)_{pq} = \max_{i \in Q}(\varphi(x)_{pi} \bullet \varphi(y)_{iq})$, $\forall p, q \in Q$, where $\bullet$ denotes multiplication $\mod 3$.

   In order to show $*$ is associative, it suffices to show $\forall x, y \in \sum^+, \varphi(xy) = \varphi(x) * \varphi(y)$.

   Let $x, y \in \sum^+$ be arbitrary, and let $p, q \in Q$.

We know $\varphi(xy)_{pq} \in \{0, 1, 2\}$.

case 1: $\varphi(xy)_{pq} = 0$. By construction, this means there is no valid path on B labeled xy from state p to state q, *i.e.* $\forall i \in Q$, $\varphi(x)_{pi} = 0$ or $\varphi(y)_{iq} = 0$.

Therefore, $\varphi^*(xy)_{pq} = \max\limits_{i \in Q}(\varphi(x)_{pi} \bullet \varphi(y)_{iq}) = 0$.

case 2: $\varphi(xy)_{pq} = 1$. By construction, this implies there exists a valid path on B from p to q that is labeled xy without containing a final state, *i.e.* $\exists i \in Q$ *s.t.* $\varphi(x)_{pi} = 1$ and $\varphi(y)_{iq} = 1$.

Therefore, $\varphi^*(xy)_{pq} = \max\limits_{i \in Q}(\varphi(x)_{pi} \bullet \varphi(y)_{iq}) = 1$.

case 3: $\varphi(xy)_{pq} = 2$. By construction, this implies there exists a valid path on B from p to q that is labeled xy and which contains a final state, *i.e.* $\exists i \in Q$ *s.t.* $\varphi(x)_{pi} = 1$ and $\varphi(y)_{iq} = 2$, or $\varphi(x)_{pi} = 2$ and $\varphi(y)_{iq} = 1$ or $\varphi(x)_{pi} = \varphi(y)_{iq} = 2$.

Therefore, $\varphi^*(xy)_{pq} = \max\limits_{i \in Q}(\varphi(x)_{pi} \bullet \varphi(y)_{iq}) = 2$.

From the three cases, we conclude $\forall p, q \in Q, \varphi(xy)_{pq} = \varphi^*(xy)_{pq}$.

Therefore, $\forall x, y \in \sum^+, \varphi(xy) = \varphi(x) * \varphi(y)$.

This is a sufficient condition for associativity since $\forall x, y, z \in \sum^+$,

$\varphi(x) * \varphi(y) * \varphi(z) = \varphi(xyz) = \varphi(xy) * \varphi(z) = \varphi(x) * \varphi(yz)$.

Hence $S = \varphi(\sum^+)$ is a semigroup.

2. Now, let us show φ strongly recognizes $X = L^\omega(B)$.

   Let $\mathcal{L}$ be the set of linked pairs over S and

   $P = \{(R, E) \in \mathcal{L} \mid \exists p, q \in I \times Q \; s.t. \; R_{pq} \geq 1 \wedge E_{qq} = 2\}$.

   We shall prove $X = \bigcup\limits_{(R,E) \in P} [R][E]^\omega = [P]$.

   $\mathbf{X} \subseteq [\mathbf{P}]$: Suppose $u \in X = L^\omega(B)$. Then, by Theorem 2.5, $\exists (R, E) \in \mathcal{L}$ and a factorization $u = u_0 u_1 ...$ over $\sum^+$ *s.t.* $\varphi(u_0) = R$

   and $\varphi(u_i) = E, \forall i > 0$.

Since B accepts u, we know there is a valid path on B labeled u that visits a final state infinitely often. Therefore, $\exists (u_{\delta_i})_{i\geq 0}$ a subsequence of $(u_i)_{i>0}$ $s.t.\ \exists p_{\delta_i}, q_{\delta_i} \in Q$ satisfying $\varphi(u_{\delta_i})_{p_{\delta_i} q_{\delta_i}} = 2$, $\forall i \geq 0$.

Since $|Q| < \infty$, $\exists i, j \in \mathbb{N}\ s.t.\ p_{\delta_i} = p_{\delta_j}$.

We know $\varphi(u_{\delta_i} ... u_{\delta_j - 1}) = \varphi(u_{\delta_i}) * ... * \varphi(u_{\delta_j - 1}) = E$.

Since $u_{\delta_i} ... u_{\delta_j - 1}$ labels a valid path from p_{δ_i} to $p_{\delta_j} = p_{\delta_i}$ and contains at least one final state, $E_{p_{\delta_i} p_{\delta_i}} = \varphi(u_{\delta_i} ... u_{\delta_j - 1})_{p_{\delta_i} p_{\delta_i}} = 2$.

Let $q = p_{\delta_i}$. Then $\exists q \in Q\ s.t.\ E_{qq} = 2$.

Let $R' = \varphi(u_0 ... u_{\delta_i - 1})$. Since $u_0 ... u_{\delta_i - 1}$ labels a valid path, $\exists p \in I \subseteq Q\ s.t.$ $R'_{pq} \geq 1$ as desired.

Note $R' = \varphi(u_0 ... u_{\delta_i - 1}) = \varphi(u_0)\varphi(u_1 ... u_{\delta_i - 1}) = R * E = R$ since $(R, E) \in \mathcal{L}$.

We conclude $u \in [R][E]^\omega\ s.t.\ (R, E) \in P$.

Hence $u \in [P]$.

Therefore $X \subseteq [P]$. (1)

$[\mathbf{P}] \subseteq \mathbf{X}$: Let $u \in \sum^\omega$.

We know $u \in [P] \implies \exists (R, E) \in P\ s.t.\ u \in [R][E]^\omega$.

Therefore, there exists a factorization $u = u_0 u_1 ...$ as a product of words in $\sum^+$ $s.t.\ \varphi(u_0) = R$, $\varphi(u_i) = E, \forall i > 0$ and $(R, E) \in P$.

We know $(R, E) \in P \implies \exists p, q \in I \times Q\ s.t.\ R_{pq} \geq 1 \wedge E_{qq} = 2$.

Then $(\varphi(u_0) = R) \wedge (R_{pq} \geq 1) \implies \exists$ a valid path $q_0 ... q_{|u_0|}\ s.t.\ q_0 = p \in I$ and $q_{|u_0|} = q$ that corresponds to the computation of u_0 on B.

Similarly, $(\varphi(u_i) = E\ , \forall i > 0) \wedge (E_{qq} = 2) \implies \exists$ a valid path $q_{10} ... q_{1|u_1|} q_{20} ... q_{2|u_2|} ...\ s.t.$ $\forall i > 0, q_{i0} = q_{i|u_i|} = q$ and each segment $q_{i0} ... q_{i|u_i|}$ contains a final state and corresponds to a computation of u_i on B.

Therefore, $q_0 ... q_{|u_0|} q_{10} ... q_{1|u_1|} q_{20} ... q_{2|u_2|} ...$ is a valid path on B that corresponds to the computation of $u = u_0u_1...$ and that visits a final state infinitely often starting at $q_0 = q \in I$.

Hence $u \in X = L^\omega(B)$.

We thereby conclude $[P] \subseteq X$. (2)

From (1) and (2), we conclude $X = [P]$, and therefore the language is strongly recognized by the morphism φ.

- **3** $\implies$ **4** : Suppose X is strongly recognized by a morphism into a finite semigroup S. We shall show X is recognized by some morphism onto a finite $\omega-$semigroup.

  In section 2.4, we presented a method to extend a semigroup to an $\omega-$semigroup. Similarly, we construct the $\omega-$semigroup $\bar{S} = (S_+, S_\omega)$ extending $S_+ = S$.

  Let us prove X is recognized by some morphism onto $\bar{S}$, *i.e.* show $\exists\, \varphi : \sum^\infty \to \bar{S}$ and a pair $P = (P_+, P_\omega)$ with $P_+ \subseteq S_+$ and $P_\omega \subseteq S_\omega$ such that $X = \varphi^{-1}(P)$.

  We know $X_+ = X \cap \sum^+ = \emptyset$ since $X \subseteq \sum^\omega$.

  Consequently, $P_+ = \varphi_+(X_+) = \emptyset \subseteq S_+$.

  Let $\mathcal{L}$ be the set of linked pairs of S_+. Since X is strongly recognized by a morphism φ_+ into S_+, $X = \bigcup\limits_{(s,e)\in T} \varphi_+^{-1}(s)(\varphi_+^{-1}(e))^\omega$ where $T \subseteq \mathcal{L}$.

  It suffices then to set for all $u \in \sum^\omega$, $\varphi_\omega(u) = (s, e)$ where (s, e) is the linked pair generating u. Consequently, $P_\omega = \{R \in S_\omega | R = (s, e) \in T\} = \varphi_\omega(X_\omega)$. Note that P_ω is well defined because S_ω is the set of conjugacy classes of linked pairs of S_+ by construction.

  Hence we get the desired result for $\varphi = (\varphi_+, \varphi_\omega)$ and $P = (P_+, P_\omega)$ as defined above.

- **4** $\implies$ **1** : Now, suppose X is recognized by some morphism onto a finite $\omega-$semigroup. We shall show X is $\omega-$rational.

Let $\bar{S} = (S_+, S_\omega)$ be the $\omega-$semigroup and $\varphi = (\varphi_+, \varphi_\omega)$ be the morphism. We know

$$
\begin{aligned}
u \in X &\implies u = u_1 u_2 ... \, s.t. \, u_i \in \Sigma^+, \forall i \in \mathbb{N}. \\
&\implies \varphi_\omega(u) = \varphi_+(u_1)\varphi_+(u_1)... \\
&\implies \varphi_\omega(u) = se^\omega \text{ for } \; u = u_{\delta_1} u_{\delta_2} ... s.t. \; (s = \varphi_+(u_{\delta_1})) \wedge (e = \varphi_+(u_{\delta_i}), \forall i > 1). \\
&\implies x = \varphi_+^{-1}(s)(\varphi_+^{-1}(e))^\omega \, s.t. \, s, e \in S_+.
\end{aligned}
$$

Therefore, we can write $X = \bigcup\limits_{(s,e) \in T} \varphi_+^{-1}(s)(\varphi_+^{-1}(e))^\omega$ for some $T \subseteq S_+^2$.
Let $Z = \{\varphi_+^{-1}(s) | \exists e \in S_+ \, s.t. \, (s, e) \in T\}$ and
$Y = \{\varphi_+^{-1}(e) | \exists s \in S_+ \, s.t. \, (s, e) \in T\}$.
Since X and Y are each the preimage of a subset of the finite semigroup S_+ under the morphism φ_+ from $\sum^+$, and φ_+ can be extended to a morphism into a monoid from the free monoid $\sum^*$, both X and Y are rational languages [10].
Hence $X = \bigcup ZY^\omega$ is ω-rational as desired.

$\square$

Chapter 4

Conclusion

This paper serves as a comprehensive introduction to the algebraic recognition of infinite words, and as a simple guide for mathematicians and computer scientists interested in the study of the ω-language and the algebraic characterization of automata. This work primarily provides a low-level proof to the equivalence of the automatic and algebraic approaches after presenting readers with the fundamental knowledge required to build the proof. Finally, this algebraic exploration of Büchi automata represents a solid basis for the algebraic characterization of other types of automata, such as Muller automata, and prophetic automata, as well as other types of input, notably nested words and infinite trees.

Bibliography

[1] Carton, Olivier, Perrin, Dominique, Pin, Jean-Éric, *Automata and semigroups recognizing infinite words,* Logic and Automata, 2008.

[2] Fleischer, Lukas, Kufleitner, Manfred *Efficient Algorithms for Morphisms over Omega-Regular Languages,* arXiv, 2015.

[3] Giannakopoulou, Dimitra, *Model Checking for Concurrent Software Architectures,* Imperial College London, UK, 1999.

[4] Godefroid, Patrice, *Software Model Checking: The VeriSoft Approach,* Formal Methods in System Design 2003.

[5] Jhala, Ranjit, Majumdar, Rupak, *Software Model Checking,* ACM Computing Surveys 2009.

[6] Mukund, Madhavan, *Finite-state Automata on Infinite Inputs,* Internal Report, SPIC Mathematical Institute, India, 1996.

[7] Perrin, Dominique, Pin, Jean-Éric, *Infinite Words: Automata, Semigroups, Logic and Games,* Elsevier, 2004.

[8] Pin, Jean-Éric, *Mathematical Foundations of Automata Theory,* version of January 17, 2022.

[9] Straubing, Howard, *Finite automata, Formal Logic, and Circuit Complexity,* Birkhauser Verlag, CHE, 1994.

[10] Weil, Pascal, *Algebraic Recognizability of Languages,* Mathematical Foundations of Computer Science 2004, 149-175.