Ambiguity Through the Lens of Measure Theory

Olivier Carton
IRIF, Université Paris Cité & CNRS, France

Abstract

In this paper, we establish a strong link between the ambiguity for finite words of a Büchi automaton and the ambiguity for infinite words of the same automaton. This link is based on measure theory. More precisely, we show that such an automaton is unambiguous, in the sense that no finite word labels two runs with the same starting state and the same ending state if and only if for each state, the set of infinite sequences labelling two runs starting from that state has measure zero. The measure used to define these negligible sets, that is sets of measure zero, can be any measure computed by a weighted automaton which is compatible with the Büchi automaton. This latter condition is very natural: the measure must only put weight on sets $wA^{\mathbb{N}}$ where w is the label of some run in the Büchi automaton.

2012 ACM Subject Classification Theory of computation → Automata over infinite objects

Keywords and phrases ambiguity, Büchi automata, measure theory

Digital Object Identifier 10.4230/LIPIcs.FSTTCS.2022.34

Related Version *Full Version*: `https://arxiv.org/abs/2011.10534`

Funding Work supported by ANR CODYS (ANR-18-CE40-0007).

1 Introduction

The relationship between deterministic and non-deterministic machines has been extensively studied since the very beginning of computer science. Despite these efforts, many questions remain wide open. This is of course true in complexity theory for questions like P versus NP but also in automata theory [15, 12]. It is for instance not known whether the simulation of non-deterministic either one-way or two-way automata by deterministic two-way automata requires an exponential blow-up of the number of states [20].

Unambiguous machines are usually defined as non-deterministic machines in which each input has at most one accepting run. They are intermediate machines in between the two extreme cases of deterministic and non-deterministic machines. The notion of ambiguity considered in the paper is slightly stronger and more structural as it does not depend on initial and final states. In the case of automata accepting finite words, non-deterministic automata can be exponentially more succinct than unambiguous automata which can be, in turn, exponentially more succinct than deterministic automata [22]. However, the inclusion problem for unambiguous automata is tractable is polynomial time [23] like for deterministic automata while the same problem for non-deterministic automata is PSPACE-complete [1, Section 10.6].

The polynomial time algorithm for the inclusion of unambiguous automata accepting finite words in [23] is based on a clever counting argument which cannot easily be adapted to infinite words. It is still unknown whether the inclusion problem for unambiguous Büchi automata can be solved in polynomial time. The problem was solved in [17] for sub-classes of Büchi automata with weak acceptance conditions and in [6] for prophetic Büchi automata introduced in [10] (see also [19, Sec. II.10]) which are strongly unambiguous. These latter results are obtained through reductions of the problem for infinite words to the problem for finite words. The main result of this paper can be seen as a first step towards a solution for all Büchi automata as it connects ambiguity for infinite words to ambiguity for finite words and thus provides a better understanding of ambiguity for infinite words.

42nd IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2022).
Editors: Anuj Dawar and Venkatesan Guruswami; Article No. 34; pp. 34:1–34:14
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

The aim of this paper is to exhibit a strong link between the ambiguity of some automaton for finite words and the ambiguity of the same automaton for infinite words. The paper is focused on strongly connected Büchi automata. Two examples given in the conclusion show that the problem is more involved for non strongly connected automata. It turns out that unambiguity for infinite words implies the unambiguity for finite words but the converse does not hold in general. This converse can however be recovered if unambiguity for infinite words is considered up to a negligible set of inputs. *Negligible* should here be understood as a set of zero measure. The measure used to characterize ambiguity must fulfill some compatibility conditions with the automaton. Some examples show these conditions cannot be avoided. Note that measures were already used to characterize maximal variable-length codes [3, Thm. 5.10] which are, in essence, a combinatorial definition of non-ambiguity.

The first step of the proof is to show that the measure of the set of accepted sequences does not increase if all states of the automaton are made final. This result is interesting by itself but it also reduces the proof of our result to automata with all states final. Since initial states are also not relevant, the problem is again reduced to automata with all states initial and final, which accept the so called shift spaces from symbolic dynamics [18]. This special case is handled using techniques from this domain like synchronizing words and Fisher covers.

This work was motivated by questions about automata with outputs also known as transducers. These transducers realize functions mapping infinite sequences to infinite sequences and the questions are focused on the long term behaviour. A natural question is the preservation of normality where normality is the property that all blocks of the same length occur with the same limiting frequency [9]. Normality was introduced by Borel to formalize the most basic form of randomness for real numbers [5]. It turns out that normality can be characterized by non-compressibility by transducers realizing one-to-one functions [2]. Since each infinite run ends in a strongly connected component, it is sufficient to study ambiguity of strongly connected automata. It is a classical result that each function realized by a transducer can be realized by a transducer whose input automaton is unambiguous [11]. The result proved in this paper shows that if all states of a strongly connected unambiguous transducer are made final the transducer remains unambiguous up to a set of measure zero. It allows us to use, for instance, the ergodic theorem for Markov chains where the function must be defined up to a set of measure zero.

The paper is organized as follows. Section 2 is devoted to basic definitions needed for the main result which is stated in Section 3. The first step of the proof is to reduce the problem to the special case of Büchi automata with all states final. This is done in Section 4. The proof of this special case is carried out in Section 5.

2 Definitions

2.1 Words, sequences and measures

Let A be a finite set of *symbols* that we refer to as the *alphabet*. We write $A^{\mathbb{N}}$ for the set of all sequences on the alphabet A and A^* for the set of all (finite) words. The length of a finite word w is denoted by $|w|$. The positions of sequences and words are numbered starting from 1. The empty word is denoted by ε. The cardinality of a finite set E is denoted by $\#E$. A *factor* of a sequence $a_1a_2a_3\cdots$ is a finite word of the form $a_ka_{k+1}\cdots a_{\ell-1}$ for integers $1 \leqslant k \leqslant \ell$ where $k = \ell$ yields the empty word ε. We let $\mathrm{fact}(X)$ denote the set of factors of a set X of sequences.

We recall here a few notions of topology. The set $A^{\mathbb{N}}$ of sequences can be endowed with a topology by the distance d which is defined as follows. The distance $d(x, y)$ of two sequences $x = a_1a_2a_3\cdots$ and $y = b_1b_2b_3\cdots$ is zero if $x = y$ and is $2^{-\min\{i:a_i\neq b_i\}}$ otherwise. The set X

is *open* if it is equal to a possibly infinite union of *cylinders*, that is, sets of the form $wA^{\mathbb{N}}$ for $w \in A^*$. It is *closed* if its complement in $A^{\mathbb{N}}$ is open. Each set X is contained in a smallest closed set $\overline{X}$ called its *closure*. The complement of $\overline{X}$ is the union of all sets $wA^{\mathbb{N}}$ which are disjoint from X.

By measure, we mean, in this paper, a *probability measure on* $A^{\mathbb{N}}$, that is, a σ-additive set function μ which assigns to each Borelian set $X \subseteq A^{\mathbb{N}}$ a real number $\mu(X)$ (called its measure) in the interval $[0,1]$ and which satisfies $\mu(A^{\mathbb{N}}) = 1$. The σ-additivity property means that if $X_0, X_1, X_2, \ldots$ is a collection of pairwise disjoint sets, then

$$\mu(\bigcup_{n \geqslant 0} X_n) = \sum_{n \geqslant 0} \mu(X_n).$$

Most sets considered in this paper are rational, and therefore, they are Borelian, (in the level $\mathbf{\Delta}^0_3$ of the Borel hierarchy) and their measure does always exist. By the Carathéodory Extension Theorem, each measure μ is fully determined by the measures of the cylinders sets, that is, sets of the form $wA^{\mathbb{N}}$ for some finite word w. In the rest of the paper, we identify a measure μ on $A^{\mathbb{N}}$ and the function which maps each finite word w to the measure $\mu(wA^{\mathbb{N}})$ of the cylinder set $wA^{\mathbb{N}}$.

A *probability measure on* A^* is a function $\mu : A^* \to [0,1]$ such that $\mu(\varepsilon) = 1$ and that the equality

$$\sum_{a \in A} \mu(wa) = \mu(w)$$

holds for each word $w \in A^*$. The simplest example of a probability measure is a *Bernoulli measure*. It is a monoid morphism from A^* to $[0,1]$ (endowed with multiplication) such that $\sum_{a \in A} \mu(a) = 1$. Among the Bernoulli measures is the *uniform measure* which maps each word $w \in A^*$ to $(\#A)^{-|w|}$. In particular, each symbol a is mapped to $\mu(a) = 1/\#A$.

By the Carathéodory Extension Theorem, a measure μ on A^* can be uniquely extended to a probability measure $\hat{\mu}$ on $A^{\mathbb{N}}$ such that $\hat{\mu}(wA^{\mathbb{N}}) = \mu(w)$ holds for each word $w \in A^*$. As already said, we use the same symbol for μ and $\hat{\mu}$. A probability measure μ is said to be *(shift) invariant* if the equality

$$\sum_{a \in A} \mu(aw) = \mu(w)$$

holds for each word $w \in A^*$. The support $\mathrm{supp}(\mu)$ of a measure μ is the set $\mathrm{supp}(\mu) = \{w \in A^* : \mu(w) > 0\}$ of finite words.

The column vector such that each of its entries is 1 is denoted by $\mathbb{1}$. A P-vector λ is called *stochastic* (respectively, *substochastic*) if its entries are non-negative and sum up to 1 (respectively, to at most 1). that is, $0 \leqslant \lambda_p \leqslant 1$ for each $p \in P$ and $\lambda\mathbb{1} = 1$ (respectively, $\lambda\mathbb{1} \leqslant 1$). A matrix M is called *stochastic* (respectively, *substochastic*) if each of its rows is stochastic (respectively, *substochastic*), that is $M\mathbb{1} = \mathbb{1}$ (respectively, $M\mathbb{1} \leqslant \mathbb{1}$). It is called *strictly substochastic* if it is substochastic but not stochastic. This means that the entries of at least one of its rows sum up to a value which is strictly smaller than 1.

In the paper, we mainly consider rational measures also known as hidden Markov measures and under other names in the literature [7]. These measures are those for which the values $\mu(w)$ for $w \in A^*$ are given by a weighted automaton [21, Chap .4] or equivalently by a (matrix) representation. A measure μ is *rational* if there is an integer m, a row $(1 \times m)$-vector π, a morphism ν from A^* into $m \times m$-matrices over real numbers and a column $(m \times 1)$-vector ρ such that the following equality holds for each word $a_1 \cdots a_k$ [4].

$$\mu(a_1 \cdots a_k) = \pi\nu(a_1 \cdots a_k)\rho = \pi\nu(a_1) \cdots \nu(a_k)\rho$$

The triple $\langle \pi, \nu, \rho \rangle$ is called a *representation* of the rational measure μ and the integer m the *dimension.* By the main result in [14], it can always be assumed that both the vector π and the matrix $\sum_{a \in A} \nu(a)$ are stochastic and that the vector ρ is the vector $\mathbb{1}$. The triple $\langle \pi, \nu, \mathbb{1} \rangle$ is then called a *stochastic representation* of μ. Note that if the representation $\langle \pi, \nu, \rho \rangle$ is stochastic, the function μ defined by $\mu(w) = \pi\nu(w)\rho$ is a measure because it satisfies the required properties. The measure μ is invariant if $\pi \sum_{a \in A} \nu(a) = \pi$. Note that rational measures with dimension 1 are the Bernoulli measures.

We give below an example of a rational measure. Consider the stochastic representation $\langle \pi, \nu, \rho \rangle$ of dimension 2 where π is the row vector $(1, 0)$, ρ is the column vector $\mathbb{1}$ and the morphism ν from $\{0, 1\}^*$ into 2×2-matrices is defined by

$$\nu(0) = \frac{1}{3} \begin{pmatrix} 1 & 1 \\ 0 & 3 \end{pmatrix} \quad \text{and} \quad \nu(1) = \frac{1}{3} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

An equivalent weighted automaton is pictured in Figure 4. The measure defined by this representation is given for each word w of length n ending with a block of $0 \leqslant k \leqslant n$ zeros by $\mu(w) = (1 + (1 + 3 + \cdots + 3^{k-1}))/3^n = (3^k + 1)/23^n$. We will see that this measure does not meet our expectations because it is not irreducible: the weighted automaton pictured in Figure 4 is not strongly connected.

2.2 Automata and ambiguity

We refer the reader to [19] for a complete introduction to automata accepting (infinite) sequences of symbols. A *(Büchi) automaton* $\mathcal{A}$ is a tuple $\langle Q, A, \Delta, I, F \rangle$ where Q is the finite state set, A the alphabet, $\Delta \subseteq Q \times A \times Q$ the transition relation, $I \subseteq Q$ the set of initial states and F is the set of final states. A transition is a tuple $\langle p, a, q \rangle$ in $Q \times A \times Q$ and it is written $p \xrightarrow{a} q$. A *finite run* in $\mathcal{A}$ is a finite sequence of consecutive transitions,

$$q_0 \xrightarrow{a_1} q_1 \xrightarrow{a_2} q_2 \cdots q_{n-1} \xrightarrow{a_n} q_n$$

Its *label* is the word $a_1 a_2 \cdots a_n$. An *infinite run* in $\mathcal{A}$ is a sequence of consecutive transitions,

$$q_0 \xrightarrow{a_1} q_1 \xrightarrow{a_2} q_2 \xrightarrow{a_3} q_3 \cdots$$

A run is *initial* if its first state q_0 is initial, that is, belongs to I. A run is called *final* if it visits infinitely often a final state. An infinite run is *accepting* if it is both initial and final. A sequence is *accepted* if it is the label of an accepting run. The set of accepted sequences is said to be *accepted* by the automaton. As usual, an automaton is *deterministic* if it has only one initial state, that is $\#I = 1$ and if $p \xrightarrow{a} q$ and $p \xrightarrow{a} q'$ are two of its transitions with the same starting state and the same label, then $q = q'$. The automaton pictured in Figure 1 accepts the set $0^*1^{\mathbb{N}}$ of sequences having some 0s and then only 1s. The leftmost automaton pictured in Figure 2 is deterministic while the middle one is not. Both accept the set of sequences having infinitely many 1s. An automaton is *trim* if each state occurs in an accepting run.

For each state q, its *future* (respectively *bi-future*) is the set $\mathsf{F}(q)$ (respectively, $\mathsf{F}_2(q)$) of sequences labelling a final run (respectively, at least two final runs) starting from q. Let $\overline{\mathsf{F}}(q)$ (respectively, $\overline{\mathsf{F}}_2(q)$) be the set of sequences labelling at least one (respectively, two) infinite run starting from q which might be final or not. Note that if the automaton is trim, $\overline{\mathsf{F}}(q)$ is indeed the topological closure of $\mathsf{F}(q)$ but that $\overline{\mathsf{F}}_2(q)$ might not be the topological closure of $\mathsf{F}_2(q)$ as shown by the automaton pictured in Figure 1. The *past* $\mathsf{P}(q)$ of a state q is the set of finite words labelling a run ending in q. For an automaton $\mathcal{A}$, we let $\mathrm{fact}(\mathcal{A})$ denote the set of finite words labelling some run in $\mathcal{A}$. Therefore $\mathrm{fact}(\mathcal{A}) = \bigcup_{q \in Q} \mathsf{P}(q)$ where Q is the state set of $\mathcal{A}$.

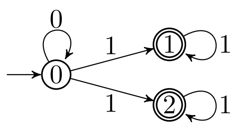


Figure 1 $\overline{\mathsf{F}}_2(0) = 0^*1^{\mathbb{N}}$ and $\overline{\mathsf{F}_2(0)} = 0^*1^{\mathbb{N}} \cup \{0^{\mathbb{N}}\}$.

An automaton is *unambiguous* (for finite words) if for each states $p, q \in Q$ and each word w, there is at most one run $p \xrightarrow{w} q$ from p to q labelled by w. Each automaton which is either deterministic or reverse-deterministic is unambiguous.

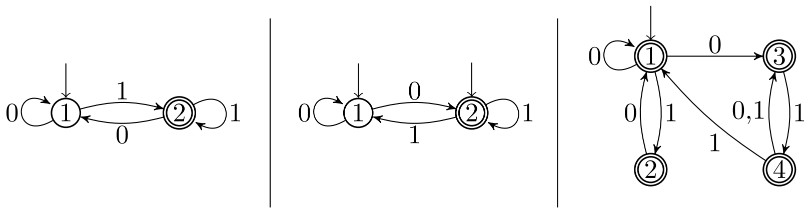


Figure 2 Three unambiguous automata.

The three automata pictured in Figure 2 are unambiguous. The leftmost one is deterministic and $\mathsf{F}(1) = \mathsf{F}(2) = (0^*1)^{\mathbb{N}}$, $\overline{\mathsf{F}}(1) = \overline{\mathsf{F}}(2) = \{0,1\}^{\mathbb{N}}$ and $\mathsf{F}_2(1) = \mathsf{F}_2(2) = \varnothing$. The middle one is reverse deterministic (that is, becomes deterministic if transitions are reversed), $\mathsf{F}(1) = 0(0^*1)^{\mathbb{N}}$, $\mathsf{F}(2) = 1(0^*1)^{\mathbb{N}}$, $\overline{\mathsf{F}}(1) = 0\{0,1\}^{\mathbb{N}}$, $\overline{\mathsf{F}}(2) = 1\{0,1\}^{\mathbb{N}}$ and $\mathsf{F}_2(1) = \mathsf{F}_2(2) = \varnothing$. The rightmost one is neither deterministic nor reverse deterministic but it is unambiguous. Note however that $\mathsf{F}_2(1)$ is not empty: $\mathsf{F}_2(1) \supset 0^*(01)^{\mathbb{N}}$. An ambiguous automaton is pictured in Figure 3 below.

With each stochastic representation $\langle \pi, \nu, \mathbb{1} \rangle$ of a rational measure is associated an automaton whose state set is $P = \{1, \ldots, m\}$ where m is the common dimension of all matrices $\nu(a)$. For each states $p, q \in P$, there is a transition $p \xrightarrow{a} q$ whenever $\nu(a)_{p,q} > 0$. The initial states are those states q in P such that $\pi_q > 0$. Due to this automaton, $\overline{\mathsf{F}}(p)$ is well-defined for a state $p \in P$. The representation is called *irreducible* if this automaton is strongly connected. A rational measure is called *irreducible* if it has at least one irreducible representation.

A strongly connected component C of graph (respectively automaton) is called *terminal* if it cannot be left, that is, if $p \to q$ is a edge with $p \in C$, then $q \in C$.

3 Main result

Ambiguity of automata has been defined using finite words: an automaton is ambiguous if some finite word w is the label of two different runs from a state p to a state q. If the automaton is trim, this implies that some sequence of the form wy is the label of two different runs from p. The converse of this implication does not hold in general. The third automaton pictured in Figure 2 is unambiguous although the sequence $(01)^{\mathbb{N}} = 0101 \cdots$ is the label of the following two accepting runs starting from state 1.

$$1 \xrightarrow{0} 1 \xrightarrow{1} 2 \xrightarrow{0} 1 \xrightarrow{1} 2 \xrightarrow{0} 1 \xrightarrow{1} \cdots$$
$$1 \xrightarrow{0} 3 \xrightarrow{1} 4 \xrightarrow{0} 3 \xrightarrow{1} 4 \xrightarrow{0} 3 \xrightarrow{1} \cdots$$

However, the set $\mathsf{F}_2(1)$ is contained in $(0+1)^*(01)^{\mathbb{N}}$ and it is thus countable and of measure 0 for the uniform measure. Note that if each transition $p \xrightarrow{1} q$ is replaced by the two transitions $p \xrightarrow{1} q$ and $p \xrightarrow{2} q$, the set $\mathsf{F}_2(1)$ is not anymore countable but it is still of measure 0 as a subset of $\{0,1,2\}^{\mathbb{N}}$.

The following theorem provides a characterization of ambiguity using measure theory. More precisely, it states that a strongly connected automaton is unambiguous whenever the measure of sequences labelling two runs is negligible, that is, of measure zero.

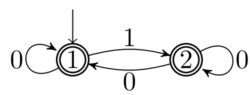
▶ **Theorem 1.** *Let $\mathcal{A}$ be a strongly connected Büchi automaton and let μ be an irreducible rational measure such that* $\mathrm{supp}(\mu) = \mathrm{fact}(\mathcal{A})$. *The following conditions are equivalent.*

i) *The automaton $\mathcal{A}$ is unambiguous.*
ii) *For each state q of $\mathcal{A}$, $\mu(\overline{\mathsf{F}}_2(q)) = 0$.*
iii) *There is a state q of $\mathcal{A}$ such that $\mu(\mathsf{F}_2(q)) = 0$.*

The equality $\mathrm{supp}(\mu) = \mathrm{fact}(\mathcal{A})$ is called the *full-support condition* in [7] because the inclusion $\mathrm{supp}(\mu) \subseteq \mathrm{fact}(\mathcal{A})$ is implicit in [7]. The irreducibility of the measure ensures that it does not put too much weight on too small sets (See example after Proposition 2). The measure used to quantify this ambiguity must also be compatible with the automaton. More precisely, its support must be equal to the set of finite words labelling at least one run in the automaton. If this condition is not fulfilled, the result may not hold as it is shown by the following two examples below.

We first explain, given a strongly connected automaton $\mathcal{A}$, how to construct an irreducible measure μ such that equality $\mathrm{supp}(\mu) = \mathrm{fact}(\mathcal{A})$ holds. Informally, the constructions consists in assigning positive weights to states and transitions of $\mathcal{A}$ such that the following two conditions are satisfied. The weights of the states must sum up to 1 (this is a distribution) and for each state q, the weights of the transitions outgoing from q must sum up to 1. Assigning weights to states is the same as defining a stochastic Q-vector π with positive entries. Assigning weights to transitions is the same as defining a $Q \times Q$-matrix $\nu(a)$ for each symbol $a \in A$ such that the matrix $\sum_{a \in A} \nu(a)$ is stochastic and such that for each states p, q and each symbol a, the (p,q) entry of $\nu(a)$ is positive if and only if $p \xrightarrow{a} q$ is a transition of $\mathcal{A}$. The measure given by the representation $\langle \pi, \nu, \mathbb{1} \rangle$ is then a rational and irreducible measure such that $\mathrm{supp}(\mu) = \mathrm{fact}(\mathcal{A})$. It is irreducible because the automaton $\mathcal{A}$ is strongly connected and each transition gets a positive weight. It satisfies $\mathrm{supp}(\mu) = \mathrm{fact}(\mathcal{A})$ because weights of states and transition are all positive. Note that not all compatible measures can be obtained that way.

Consider again the third automaton pictured in Figure 2. Let μ be the probability measure putting weight 1/2 on each of the sequences $(01)^{\mathbb{N}}$ and $(10)^{\mathbb{N}}$ and zero everywhere else. More formally, it is defined $\mu((01)^{\mathbb{N}}) = \mu((10)^{\mathbb{N}}) = 1/2$ and $\mu(\{0,1\}^{\mathbb{N}} \setminus \{(01)^{\mathbb{N}}, (10)^{\mathbb{N}}\}) = 0$. The measure $\mu(\mathsf{F}_2(1)) = 1/2$ is non-zero although the automaton is unambiguous because the support $(01)^* + (10)^*$ of this measure μ is strictly contained in the set of words labelling a run in this automaton. This latter set is actually the set $\{0,1\}^*$ of all finite words over $\{0,1\}$.



■ **Figure 3** An ambiguous automaton accepting $(0+10)^{\mathbb{N}}$.

Consider the automaton pictured in Figure 3. It accepts the set X of sequences with no consecutive 1s. It is ambiguous because the word 00 is the label of the two runs $2 \xrightarrow{0} 1 \xrightarrow{0} 1$ and $2 \xrightarrow{0} 2 \xrightarrow{0} 1$. The uniform measure $\mu(X)$ is zero. Therefore, both numbers $\mu(\mathsf{F}_2(1))$ and $\mu(\mathsf{F}_2(2))$ are zero although the automaton is ambiguous. This comes from the fact that the support $\{0,1\}^*$ of the uniform measure strictly contains the set $\mathrm{fact}(X)$. This latter set is the set $(0+10)^*(1+\varepsilon)$ of finite words with no consecutive 1s.

4 Reduction to closed sets

The purpose of this section is to show that the measure $\mu(X)$ of a rational set of sequences is closely related to the measure $\mu(\overline{X})$ of its closure as long as the measure μ is compatible with X. The main result of this section is the following proposition which is used in the proof of Theorem 1. The rest of the section is devoted to the proof of the proposition.

▶ **Proposition 2.** *Let $\mathcal{A}$ be a strongly connected Büchi automaton, q be a state of $\mathcal{A}$ and w be a word in $\mathsf{P}(q)$. Let μ be an irreducible rational measure such that* $\mathrm{supp}(\mu) = \mathrm{fact}(\mathcal{A})$. *Then $\mu(w\mathsf{F}(q)) = \mu(w\overline{\mathsf{F}}(q))$.*

The following example shows that the irreducibility assumption of the measure is indeed necessary. Consider the measure given by the weighted automaton pictured in Figure 4.

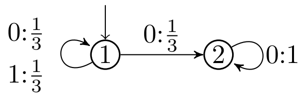


Figure 4 A weighted automaton defining a non-irreducible measure.

It realizes the measure already considered at the end of Section 2.1. This measure μ is equivalently defined by $\mu(w) = (1,0)\nu(w)\mathbb{1}$ for each finite word w where the morphism ν from $\{0,1\}^*$ into 2×2-matrices is given by

$$\nu(0) = \frac{1}{3}\begin{pmatrix} 1 & 1 \\ 0 & 3 \end{pmatrix} \quad \text{and} \quad \nu(1) = \frac{1}{3}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

The weight of a word $w1$ ending with a 1 is $3^{-|w|-1}$. Therefore, the sum of these weights when w ranges over all words of length k over the alphabet $\{0,1\}$ is given by

$$\sum_{|w|=k} \mu(w1) = \frac{2^k}{3^{k+1}} \quad \text{and} \quad \sum_{|w|\geqslant n} \mu(w1) = \sum_{k \geqslant n} \frac{2^k}{3^{k+1}} = \frac{2^n}{3^n}$$

Let $X = (0^*1)^{\mathbb{N}}$ be the set of sequences having infinitely many occurrences of the symbol 1. This set is equal to $\mathsf{F}(1)$ in the leftmost automaton pictured in Figure 2. Since X is contained in the union $\bigcup_{|w|\geqslant n} w1\{0,1\}^{\mathbb{N}}$ for each integer $n \geqslant 0$, the measure of X satisfies $\mu(X) \leqslant 2^n/3^n$ for each integer $n \geqslant 0$. This proves that the measure of X is 0 although the measure of its closure $\overline{X} = \{0,1\}^{\mathbb{N}}$ is 1.

If $\mathcal{A}$ is an automaton and $P \subseteq Q$ is a subset of its state set Q, we let $P \cdot w$ denote the subset $P' \subseteq Q$ defined by $P' = \{q : \exists p \in P \;\; p \xrightarrow{w} q\}$. If P is a singleton set $\{q\}$, we write $q \cdot w$ for $\{q\} \cdot w$. By a slight abuse of notation, we also write $q \cdot w = p$ for $\{q\} \cdot w = \{p\}$. If $\mathcal{A}$ is deterministic, $q \cdot w$ is either the empty set or a singleton set.

The following easy lemma states that, in each strongly connected automaton, there exists a run using all transitions.

▶ **Lemma 3.** *Let $\mathcal{A}$ be a strongly connected deterministic automaton. There exists a finite word w such that:*

i) *there exists a state q such that $q \cdot w$ is non-empty,*

ii) *for each state q of $\mathcal{A}$, if $q \cdot w$ is non-empty, each transition of $\mathcal{A}$ occurs in the run $q \xrightarrow{w} q \cdot w$.*

Let $\langle \pi, \nu, \mathbb{1} \rangle$ be the representation of a rational measure μ. Its support $\mathrm{supp}(\nu)$ is defined by $\mathrm{supp}(\nu) = \{w : \nu(w) \neq 0\}$. It obviously satisfies $\mathrm{supp}(\mu) \subseteq \mathrm{supp}(\nu)$. This inclusion can be strict as shown by the following example but it becomes an equality as soon as $\mathrm{supp}(\mu)$ is *factorial*, that is closed under taking factor.

Let μ be the measure defined by $\mu(0w) = 0$ and $\mu(1w) = 2^{-|w|}$ for each word w in $\{0,1\}^*$. It is rational because it is defined by $\mu(w) = (0,1)\nu(w)\left(\begin{smallmatrix}1\\1\end{smallmatrix}\right)$ where the morphism ν from $\{0,1\}^*$ into 2×2-matrices is given by $\nu(0) = \frac{1}{2}\left(\begin{smallmatrix}1&1\\0&0\end{smallmatrix}\right)$ and $\nu(1) = \frac{1}{2}\left(\begin{smallmatrix}0&0\\1&1\end{smallmatrix}\right)$. The support of this measure μ is $\mathrm{supp}(\mu) = 1\{0,1\}^*$ but the support of the morphism ν is $\mathrm{supp}(\nu) = \{w \in \{0,1\}^* : \nu(w) \neq 0\}$ is $\{0,1\}^*$. The support of a rational measure and the support of one of its representations might not coincide in general but they do coincide as soon as $\mathrm{supp}(\mu)$ is factorial as stated by the following lemma.

▶ **Lemma 4.** *Let μ be an irreducible rational measure and let $\langle \pi, \nu, \mathbb{1} \rangle$ be an irreducible representation of μ. Then $\mathrm{supp}(\mu)$ is factorial if and only if the equality $\mathrm{supp}(\mu) = \mathrm{supp}(\nu)$ holds.*

In the rest of the paper, the support of each measure is a factorial set and both supports coincide.

The following lemma states the of sequences having finitely many occurrence of some finite word has zero measure. The result is well-known when the measure is a Markov measure and the word is just a symbol [8, Thm 3.3]. The irreducibility and rationality of the measure are both crucial.

▶ **Lemma 5.** *Let μ be an irreducible rational measure such that $\mathrm{supp}(\mu)$ is factorial and let w be a word in $\mathrm{supp}(\mu)$. Then*

$$\mu(\{x : |x|_w < \infty\}) = 0.$$

where $|x|_w$ is the number of occurrences of w in x.

We let $\xrightarrow{*}$ denote the accessibility relation in an automaton. We write $p \xrightarrow{*} q$ if there is a run from p to q. If P and P' are two subsets of states of an automaton, we write $P \xrightarrow{*} P'$ whenever there is a run from a state in P to a state in P'. This relation is not transitive in general but it is when each considered subset is contained in a strongly connected component. The following lemma gives a property of Muller automata accepting the same set as a strongly connected Büchi automaton.

▶ **Lemma 6.** *Let $X \subseteq A^{\mathbb{N}}$ be a non-empty set of sequences accepted by a strongly connected Büchi automaton and let $w \in A^*$ be a finite word. Let $\mathcal{A}$ be a Muller automaton accepting the set wX and let $\mathcal{T}$ be its table of accepting subsets of states. Let F be an element of $\mathcal{T}$ such that $F' \in \mathcal{T}$ and $F \xrightarrow{*} F'$ imply $F' \xrightarrow{*} F$. Then the strongly connected component containing F also belongs to the table $\mathcal{T}$.*

A subset P of states of a Muller automaton $\mathcal{A}$ is called *essential* if there is an infinite run in $\mathcal{A}$ such that P is the set of states that occur infinitely often along this run.

Proof of Proposition 2. The proof of the proposition is reduced to proving that $\mu(w\overline{\mathsf{F}}(q) \setminus w\mathsf{F}(q)) = 0$. We consider a Muller automaton accepting $w\mathsf{F}(q)$ with a table $\mathcal{T}$. Note that a Muller automaton accepting $w\overline{\mathsf{F}}(q)$ is obtained by replacing the table $\mathcal{T}$ by the table $\overline{\mathcal{T}}$ which contains each essential set of states which can access an essential set of states in $\mathcal{T}$. The difference set $w\overline{\mathsf{F}}(q) \setminus w\mathsf{F}(q)$ is thus accepted by the same Muller automaton with the table $\mathcal{T} \setminus \overline{\mathcal{T}}$. By Lemma 6, each maximal essential state is in the table $\mathcal{T}$. By combining Lemmas 3 and 5, it is clear that $\mu(w\overline{\mathsf{F}}(q) \setminus w\mathsf{F}(q)) = 0$. ◀

5 Proof for closed sets

Thanks to Proposition 2, it is sufficient to study closed sets. As the initial states of the automaton are not relevant for the statement of Theorem 1, we consider automata where all states are initial and final, that is, $I = F = Q$. It turns out that these automata accept shift spaces that we now introduce.

The *shift map* is the function σ which maps each sequence $(x_i)_{i\geqslant 1}$ to the sequence $(x_{i+1})_{i\geqslant 1}$ obtained by removing its first element. A *shift space* is a subset X of $A^{\mathbb{N}}$ which is closed for the usual product topology and such that $\sigma(X) = X$. A classical example of a shift space is the *golden mean shift*: it is the set $\{0, 10\}^{\mathbb{N}}$ of sequences with no consecutive 1s. We refer the reader to [18] for a complete introduction to shift spaces.

If a shift space is accepted by some trim Büchi automaton, it is also accepted by the same automaton in which each state is made initial and final. A shift space is called *sofic* if it is accepted by some automaton. A sofic shift is called *irreducible* if it is accepted by a strongly connected Büchi automaton. It is well-known [18] that each shift space is characterized by the set of factors of its sequences. Let us recall that $\mathrm{fact}(X)$ denotes the set of factors of a shift space X.

There is a unique, up to isomorphism, deterministic automaton accepting an irreducible sofic shift with the minimal number of states [18, Thm 3.3.18]. This *minimal automaton* is also referred to as either its Shannon cover or its Fischer cover. It can be obtained from any automaton accepting the shift space via determinizing and state-minimizing algorithms, e.g., [18, pp. 92], [16, pp. 68]. The minimal automaton of the golden mean shift is the leftmost automaton pictured in Figure 5. A *synchronizing word* of a strongly connected automaton is a word w such that there is a unique state q such that $w \in \mathsf{P}(q)$. The word 1 is a synchronizing word of both automata pictured in Figure 5. The minimal automaton of a sofic shift has always at least one synchronizing word [18, Prop. 3.3.16].

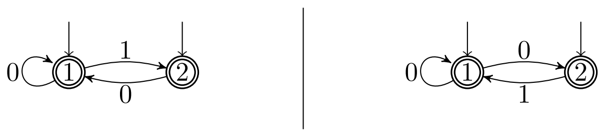


Figure 5 Two automata accepting the golden mean shift.

The next lemma states in particular (for $w = \varepsilon$) that the future of each state of the minimal automaton of a sofic shift has a positive measure. The

▶ **Lemma 7.** *Let r be a state of the minimal automaton of an irreducible sofic shift space X and let w be a synchronizing word such that $w \in \mathsf{P}(r)$. Let μ be a rational measure such that* $\mathrm{supp}(\mu) = \mathrm{fact}(X)$. *Then $\mu(w\overline{\mathsf{F}}(r)) = \mu(wA^{\mathbb{N}}) > 0$.*

The following lemma establishes a link between any automaton accepting a sofic system and its minimal automaton. It allows us to transfer the result of the previous lemma to non-minimal automata.

▶ **Lemma 8.** *Let $\mathcal{A}$ be a strongly connected automaton accepting an irreducible sofic shift space X. Let w be a word and q be a state of $\mathcal{A}$ such that $w \in \mathsf{P}(q)$. There exists a state r of the minimal automaton of X and a synchronizing word v of this minimal automaton such that $wv \in \mathsf{P}(r)$ and $\overline{\mathsf{F}}(q) \cap vA^{\mathbb{N}} = v\overline{\mathsf{F}}(r)$.*

Combining Lemmas 7 and 8 yields the following result.

▶ **Lemma 9.** *Let $\mathcal{A}$ be an strongly connected automaton accepting an irreducible sofic shift X. Let q be a state of $\mathcal{A}$ and let w be a word such that $w \in \mathsf{P}(q)$. Let μ be an irreducible rational measure such that* $\mathrm{supp}(\mu) = \mathrm{fact}(X)$. *Then $\mu(w\overline{\mathsf{F}}(q)) > 0$.*

Proof. By Lemma 8, there exists a state r of the minimal automaton of X and a synchronizing word v such that $wv \in \mathsf{P}(r)$ and $\overline{\mathsf{F}}(q) \cap vA^{\mathbb{N}} = v\overline{\mathsf{F}}(r)$. This implies that $w\overline{\mathsf{F}}(q) \cap wvA^{\mathbb{N}} = wv\overline{\mathsf{F}}(r)$. By Lemma 7, the measure $\mu(wv\overline{\mathsf{F}}(r))$ is positive and thus $\mu(w\overline{\mathsf{F}}(q)) > 0$. ◀

It is a very classical result that not all regular sets of sequences are accepted by deterministic Büchi automata. This is the reason why Muller automata with a more involved acceptance condition were introduced. Landweber's theorem states that a regular set of sequences is accepted by a deterministic Büchi automaton if and only it is a G_δ-set (that is Π^2_0) [19, Thm I.9.9]. This implies in particular that regular and closed sets[1] are accepted by deterministic Büchi automata. Regular and closed sets are actually accepted by deterministic Büchi automata in which each state is final [19, Prop III.3.7].

Lemma 9 states that the future of a state in automaton with all states final has a positive measure. The following provides a converse. It states that a closed set F with positive measure contains the future of a state of the minimal automaton, prefixed by some word w. The prefix w is really needed because the closed set F can be arbitrarily small.

▶ **Lemma 10.** *Let X be a sofic shift space and let μ be an irreducible rational measure such that* $\mathrm{supp}(\mu) = \mathrm{fact}(X)$. *Let F be a regular and closed set contained in X. If $\mu(F) > 0$, there exists a word w and a state r of the minimal automaton of X such that $w \in \mathsf{P}(r)$ and $w\overline{\mathsf{F}}(r) \subseteq F$.*

Before proceeding to the proof of the lemma, we show that even in the case of the full shift, that is $X = A^{\mathbb{N}}$, both hypothesis of being regular and closed are necessary. Since the minimal automaton of the full shift has a single state r satisfying $\overline{\mathsf{F}}(r) = A^{\mathbb{N}}$, the lemma can be, in that case, rephrased as follows. If $\mu(F) > 0$ where μ is the uniform measure, then there exists a word w such that $wA^{\mathbb{N}} \subseteq F$.

Being regular is of course not sufficient because the set $(0^*1)^{\mathbb{N}}$ of sequences having infinitely many occurrences of 1 is regular and has measure 1 but does not contain any set of the form $wA^{\mathbb{N}}$. Being closed is also not sufficient as it is shown by the following example. Let X be the set of sequences such that none of their non-empty prefixes of even length is a palindrome. The complement of X is equal to the following union

$$\bigcup_{n \geqslant 1} Z_n \quad \text{where} \quad Z_n = \bigcup_{|w|=n} w\tilde{w}A^{\mathbb{N}}$$

and where $\tilde{w}$ stands for the reverse of w. Suppose for instance that the alphabet is $A = \{0, 1\}$. The measure of Z_n is equal 2^{-n} because there are 2^n words of length n and the measure of each cylinder $w\tilde{w}A^{\mathbb{N}}$ is 2^{-2n}. Furthermore, the set $Z_1 \cup Z_2$ is equal to

[1] Not to be confused with *regular closed sets* which are equal to the closure of their interior [13, Chap. 4].

$00A^{\mathbb{N}} \cup 11A^{\mathbb{N}} \cup 0110A^{\mathbb{N}} \cup 1001A^{\mathbb{N}}$ whose measure is $5/8$. This shows that the measure of the complement of X is bounded by $5/8 + \sum_{n \geqslant 3} 2^{-n} = 7/8$ (Note that this is really an upper bound as the sets Z_n are not pairwise disjoint). Therefore X has a positive measure but it does not contain any cylinder. Indeed, in each cylinder $wA^{\mathbb{N}}$, the cylinder $w\tilde{w}A^{\mathbb{N}}$ is out of X.

The following result is trivially true when the measure μ is shift invariant because $\mu(F) = \sum_{|w|=m} \mu(wF)$ but it does not hold in general.

▶ **Lemma 11.** *Let X be a sofic shift space and let μ be an irreducible rational measure such that* $\operatorname{supp}(\mu) = \operatorname{fact}(X)$. *Let F be a regular and closed set contained in X. If $\mu(F) = 0$, then $\mu(wF) = 0$ for each finite word w.*

Proof. We prove that $\mu(wF) > 0$ implies $\mu(F) > 0$. Suppose that $\mu(wF) > 0$. Since wF is also regular and closed, there exists, by Lemma 10, a word u and a state r of the minimal automaton of X such that $u \in \mathsf{P}(r)$ and $u\overline{\mathsf{F}}(r) \subseteq wF$. This latter inclusion implies that either u is a prefix of w or w is a prefix of u. In the first case, that is $w = uv$ for some word v, the inclusion is equivalent to $\overline{\mathsf{F}}(r) \subseteq vF$. Let s be state such that $r \xrightarrow{v} s$. Then $v\overline{\mathsf{F}}(s)$ is contained in $\overline{\mathsf{F}}(r)$ and thus $\overline{\mathsf{F}}(s) \subseteq F$. By Lemma 9, $\mu(\overline{\mathsf{F}}(s)) > 0$ and thus $\mu(F) > 0$. In the second case, that is $u = wv$, for some v, the inclusion is equivalent to $v\overline{\mathsf{F}}(r) \subseteq F$. Again by Lemma 9, $\mu(v\overline{\mathsf{F}}(r)) > 0$ and thus $\mu(F) > 0$. ◀

The following lemma is an extension to pairs of futures of states of the result of Lemma 7 for one state.

▶ **Lemma 12.** *Let $\mathcal{A}$ be strongly connected automaton accepting a shift space X. Let μ be an irreducible rational measure such that* $\operatorname{supp}(\mu) = \operatorname{fact}(X)$. *Let q and q' two states of $\mathcal{A}$ such that $\mu(\overline{\mathsf{F}}(q) \cap \overline{\mathsf{F}}(q')) > 0$. Then there exists a word w such that $\overline{\mathsf{F}}(q) \cap wA^{\mathbb{N}} = \overline{\mathsf{F}}(q') \cap wA^{\mathbb{N}}$.*

Proof. We claim that there exists a word w and a state r of the minimal automaton of X such that $w \in \mathsf{P}(r)$ and

$$\overline{\mathsf{F}}(q) \cap wA^{\mathbb{N}} = \overline{\mathsf{F}}(q') \cap wA^{\mathbb{N}} = w\overline{\mathsf{F}}(r).$$

Let F be the closed set $\overline{\mathsf{F}}(q) \cap \overline{\mathsf{F}}(q')$. By Lemma 10 applied to F, there exists a word u and a state s of the minimal automaton of X such that $u \in \mathsf{P}(s)$ and

$$u\overline{\mathsf{F}}(s) \subseteq \overline{\mathsf{F}}(q)$$
$$u\overline{\mathsf{F}}(s) \subseteq \overline{\mathsf{F}}(q')$$

Let v be a synchronizing word of the minimal automaton of X such that $s \cdot v$ is not empty. Let w be the word uv and let r be the state $s \cdot v$. Since $u \in \mathsf{P}(s)$ and $r = s \cdot v$, $w \in \mathsf{P}(r)$. We claim that $\overline{\mathsf{F}}(q) \cap wA^{\mathbb{N}} = w\overline{\mathsf{F}}(r)$. Suppose first that x belongs to $\overline{\mathsf{F}}(q) \cap wA^{\mathbb{N}}$. The sequence x is then equal to wx' for some sequence x' and it is the label of a run in the minimal automaton of X. Since $w = uv$ and v is synchronizing, the sequence x' must belong to $\overline{\mathsf{F}}(r)$. Suppose conversely that x belongs to $w\overline{\mathsf{F}}(r)$. It is then equal to uvx' for some x' in $\overline{\mathsf{F}}(r)$. Since $r = s \cdot v$, $vx' \in \overline{\mathsf{F}}(s)$. It follows from the inclusion $u\overline{\mathsf{F}}(s) \subseteq \overline{\mathsf{F}}(q)$ that x belongs to $\overline{\mathsf{F}}(q)$. This completes the proof of the equality $\overline{\mathsf{F}}(q) \cap wA^{\mathbb{N}} = w\overline{\mathsf{F}}(r)$. By symmetry, the equality $\overline{\mathsf{F}}(q') \cap wA^{\mathbb{N}} = w\overline{\mathsf{F}}(r)$ also holds and the proof is completed. ◀

This last lemma establishes a link between unambiguity of an automaton and measures of futures of its states. More precisely, it states that the future of two states that can be reached from the same state and reading the same word have an intersection of zero measure. Its proof is more combinatorial than previous ones.

▶ **Lemma 13.** *Let $\mathcal{A}$ be an unambiguous strongly connected automaton accepting a shift space X. Let μ be an irreducible rational measure such that* $\mathrm{supp}(\mu) = \mathrm{fact}(X)$*. If there are two runs $p \xrightarrow{u} q$ and $p \xrightarrow{u} q'$, with $q \neq q'$, then $\mu(\overline{\mathsf{F}}(q) \cap \overline{\mathsf{F}}(q')) = 0$.*

Proof. Suppose by contradiction that $\mu(\overline{\mathsf{F}}(q) \cap \overline{\mathsf{F}}(q')) > 0$. There exists, by Lemma 12, a word v such that $\overline{\mathsf{F}}(q) \cap vA^{\mathbb{N}} = \overline{\mathsf{F}}(q') \cap vA^{\mathbb{N}}$. Let $q \cdot v$ (respectively $q' \cdot v$) be the set $\{q_1, \ldots, q_r\}$ (respectively $\{q'_1, \ldots, q'_{r'}\}$). Since $\overline{\mathsf{F}}(q) \cap vA^{\mathbb{N}} = \overline{\mathsf{F}}(q') \cap vA^{\mathbb{N}}$, the equality $\overline{\mathsf{F}}(q_1) \cup \cdots \cup \overline{\mathsf{F}}(q_r) = \overline{\mathsf{F}}(q'_1) \cup \cdots \cup \overline{\mathsf{F}}(q'_{r'})$ holds. Since the automaton is strongly connected, there is a run $q_1 \xrightarrow{w} p$ from q_1 to p. Combining this run with the run $p \xrightarrow{u} q \xrightarrow{v} q_1$ yields the cyclic run $q_1 \xrightarrow{wuv} q_1$. Since $\overline{\mathsf{F}}(q_1) \cup \cdots \cup \overline{\mathsf{F}}(q_r) = \overline{\mathsf{F}}(q'_1) \cup \cdots \cup \overline{\mathsf{F}}(q'_{r'})$, the sequence $(wuv)^{\mathbb{N}} = wuvwuv \cdots$ belongs to a set $\overline{\mathsf{F}}(q'_i)$ for some $1 \leqslant i \leqslant r'$. By symmetry, it can be assumed that $(wuv)^{\mathbb{N}} \in \overline{\mathsf{F}}(q'_1)$. There exists then a run starting from q'_1 with label $(wuv)^{\mathbb{N}}$. This run can be decomposed

$$q'_1 \xrightarrow{wuv} p_1 \xrightarrow{wuv} p_2 \xrightarrow{wuv} p_3 \cdots .$$

Since there are finitely many states, there are two integers $k, \ell \geqslant 1$ such that $p_k = p_{k+\ell}$. There are then the following two runs from p to $p_k = p_{k+\ell}$ with the same label $(uvw)^{k+\ell}uv$.

$$p \xrightarrow{u} q \xrightarrow{v} q_1 \xrightarrow{(wuv)^{k-1}} q_1 \xrightarrow{w} p \xrightarrow{uv} q'_1 \xrightarrow{(wuv)^{\ell}} p_k$$
$$p \xrightarrow{u} q' \xrightarrow{v} q'_1 \xrightarrow{(wuv)^{k+\ell}} p_{k+\ell}$$

This is a contradiction with the fact that $\mathcal{A}$ is unambiguous. ◀

Proof of Theorem 1. We first prove that (i) implies (ii). suppose that the automaton $\mathcal{A}$ is unambiguous. We show that $\mu(\overline{\mathsf{F}}_2(p)) = 0$ for each state p. We start by a decomposition of the set $\overline{\mathsf{F}}_2(p)$. Let $x = a_1a_2a_3 \cdots$ be a sequence in $\overline{\mathsf{F}}_2(p)$ and let ρ and ρ' be the two different runs labelled by x. Suppose that

$$\rho = q_0 \xrightarrow{a_1} q_1 \xrightarrow{a_2} q_2 \xrightarrow{a_3} q_3 \cdots$$
$$\rho' = q'_0 \xrightarrow{a_1} q'_1 \xrightarrow{a_2} q'_2 \xrightarrow{a_3} q'_3 \cdots$$

where $q_0 = q'_0 = p$. Let n be the least integer such that $q_n \neq q'_n$. Let a be the symbol a_n, w be the prefix $a_1 \cdots a_{n-1}$ and x' be the tail $a_{n+1}a_{n+2}a_{n+3} \cdots$. The sequence x is equal to wax' and there is a finite run $q_0 \xrightarrow{w} q_{n-1}$, two transitions $q_{n-1} \xrightarrow{a} q_n$ and $q_{n-1} \xrightarrow{a} q'_n$, and the tail x' belongs to the intersection $\overline{\mathsf{F}}(q_n) \cap \overline{\mathsf{F}}(q'_n)$. We have actually proved the following equality expressing $\overline{\mathsf{F}}_2(p)$ in term of a union of intersections of sets $\overline{\mathsf{F}}(q)$.

$$\overline{\mathsf{F}}_2(p) = \bigcup_{\substack{p \xrightarrow{w} p' \\ p' \xrightarrow{a} q \\ p' \xrightarrow{a} q'}} wa(\overline{\mathsf{F}}(q) \cap \overline{\mathsf{F}}(q'))$$

Since the union is countable, it suffices to prove that if there are two transitions $p \xrightarrow{a} q$ and $p \xrightarrow{a} q'$ with $q \neq q'$, then $\mu(\overline{\mathsf{F}}(q) \cap \overline{\mathsf{F}}(q')) = 0$. Lemma 13 and Lemma 11 allow us to conclude.

The fact that (ii) implies (iii) is clear because the set $\mathsf{F}_2(q)$ is contained in $\overline{\mathsf{F}}_2(q)$ for each state q of $\mathcal{A}$.

We now prove that (iii) implies (i). Let q be state of $\mathcal{A}$ and suppose that there are two different runs from state p to state r with the same label w. Let v the label of a run from q to p. This shows that $vw\mathsf{F}(r) \subseteq \mathsf{F}_2(q)$. Since $vw \in \mathsf{P}(r)$, the measure $\mu(vw\mathsf{F}(r))$ satisfies $\mu(vw\mathsf{F}(r)) = \mu(vw\overline{\mathsf{F}}(r))$ by Proposition 2. The measure $\mu(vw\overline{\mathsf{F}}(r))$ satisfies $\mu(vw\overline{\mathsf{F}}(r)) > 0$ by Lemma 10 and thus $\mu(\mathsf{F}_2(q)) > 0$. This completes the proof of this implication. ◀

Conclusion

As a conclusion, we would like the emphasize the difficulty of extending the result to non strongly connected automata. Consider the two automata pictured in Figure 6. The leftmost one in unambiguous where as the rightmost one is obviously ambiguous for finite words. However, $\mathsf{F}_2(0) = \overline{\mathsf{F}}_2(0) = 0^*1^{\mathbb{N}}$ and $\mathsf{F}_2(q) = \varnothing$ hold for $q \neq 0$ in both automata. The only way to distinguish one automaton from the other one is to have two different measures. In order to have $\mu_1(\mathsf{F}_2(0)) = 0$ for the leftmost automaton, the measure μ_1 should put all the weight on $0^{\mathbb{N}}$: $\mu_1(0^{\mathbb{N}}) = 1$. In order to have $\mu_2(\mathsf{F}_2(0)) > 0$ for the rightmost automaton, the measure μ_2 should put some weight on a sequence $0^n1^{\mathbb{N}}$ for some integer $n \geqslant 0$. It is not clear why the measures should be different because the set 0^*1^* of finite words labelling some run is the same in both automata.

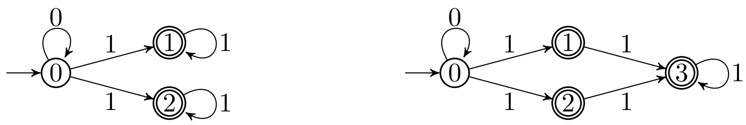


Figure 6 Two non strongly connected automata.

References

1 A. V. Aho, J. E. Hopcroft, and J. D. Ullman. *The Design and Analysis of Computer Algorithms*. Addison-Wesley, 1974.

2 V. Becher and O. Carton. Normal numbers and computer science. In V. Berthé and M. Rigo, editors, *Sequences, Groups, and Number Theory*, Trends in Mathematics Series, pages 233–269. Birkhäuser, 2018.

3 J. Berstel and D. Perrin. *Theory of Codes*. Academic Press, 1984.

4 J. Berstel and Ch. Reutenauer. *Noncommutative Rational Series with Applications*. Cambridge University Press, 2010.

5 É. Borel. Les probabilités dénombrables et leurs applications arithmétiques. *Rendiconti Circ. Mat. Palermo*, 27:247–271, 1909.

6 N. Bousquet and Ch. Löding. Equivalence and inclusion problem for strongly unambiguous Büchi automata. In *LATA 2010*, volume 6031 of *Lecture Notes in Computer Science*, pages 118–129. Springer, 2010. `doi:10.1007/978-3-642-13089-2_10`.

7 M. Boyle and K. Petersen. *Entropy of Hidden Markov Processes and Connections to Dynamical Systems*, volume 385 of *London Mathematical Society Lecture Notes*, chapter Hidden Markov processes in the context of symbolic dynamics, pages 5–71. Cambridge University Press, 2011.

8 P. Brémaud. *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*. Springer, 2008.

9 O. Carton. Preservation of normality by unambiguous transducers. *CoRR*, abs/2006.00891, 2022. `arXiv:2006.00891`.

10 O. Carton and M. Michel. Unambiguous Büchi automata. *Theoret. Comput. Sci.*, 297:37–81, 2003.

11 Ch. Choffrut and S. Grigorieff. Uniformization of rational relations. In *Jewels are Forever, Contributions on Theoretical Computer Science in Honor of Arto Salomaa*, pages 59–71. Springer, 1999.

12 Th. Colcombet. Unambiguity in automata theory. In *Descriptional Complexity of Formal Systems*, volume 9118 of *Lecture Notes in Computer Science*, pages 3–18. Springer, 2015.

13 P. R. Halmos. *Lectures on Boolean algebras*. Von Nostrand, 1963.

14 G. Hansel and D. Perrin. Mesures de probabilité rationnelles. In M. Lothaire, editor, *Mots*, pages 335–357. Hermes, 1990.

15 M. Holzer and M. Kutrib. Descriptional complexity of (un)ambiguous finite state machines and pushdown automata. In *Reachability Problems*, pages 1–23. Springer, 2010. doi:10.1007/978-3-642-15349-5_1.

16 J. E. Hopcroft and J. D. Ullman. *Introduction to Automata Theory, Languages, and Computation.* Addison-Wesley, 1979.

17 D. Isaak and Ch. Löding. Efficient inclusion testing for simple classes of unambiguous ω-automata. *Inf. Process. Lett.*, 112(14-15):578–582, 2012.

18 D. Lind and B. Marcus. *An Introduction to Symbolic Dynamics and Coding.* Cambridge University Press, 1995.

19 D. Perrin and J.-É. Pin. *Infinite Words.* Elsevier, 2004.

20 G. Pighizzini. Two-way finite automata: Old and recent results. *Electronic Proceedings in Theoretical Computer Science*, 90:3–20, 2012. doi:10.4204/eptcs.90.1.

21 J. Sakarovitch. *Elements of Automata Theory.* Cambridge University Press, 2009.

22 E. Schmidt. Succinctness of descriptions of context-free, regular, and finite languages. *DAIMI Report Series*, 7(84), 1978. doi:10.7146/dpb.v7i84.6500.

23 R. E. Stearns and H. B. Hunt III. On the equivalence and containment problems for unambiguous regular expressions, regular grammars and finite automata. *SIAM J. Comput.*, 14(3):598–611, 1985. doi:10.1137/0214044.