Ambiguity in symbolic dynamics

Dominique Perrin
(joint work with Marie-Pierre Béal, and Antonio Restivo)

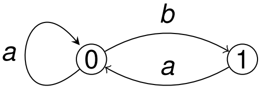
July 8, 2021

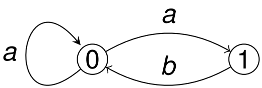
Outline

- Survey on the notion of (un)ambiguity in symbolic dynamics relating notions such as
  - unambigous automata and finite-to-one maps.
  - recognizability of morphisms
  - synchronizing automata
- Emphasis on coded systems (generalizing irreducible sofic shifts)
- Proof of an unpublished result due to Doris Fiebig.

Unambiguous automata

An *automaton* on the alphabet A is a graph with edges labeled by A. It is *unambiguous* if for every pair p, q of vertices and every word w on A, there is at most one path from p to q labeled w. It is *deterministic* if for every vertex p and letter a, there is at most one edge going out of p labeled a. Every deterministic automaton is unambiguous.





The automaton on the left is determinisic. The one on the right is unambiguous but not deterministic.

Strongly unambiguous automata

An automaton is *strongly unambiguous* if for every sequence $x \in A^{\mathbb{Z}}$, there is at most one path with label x.

The first automaton is strongly unambiguous, the second is not (there are two paths labeled $a^{\mathbb{Z}}$).

Automata and morphisms

The automaton $\mathscr{A}(\varphi)$ associated to a morphism $\varphi : A^* \to B^*$ is a bouquet of circles labeled $\varphi(a)$ for $a \in A$. It has vertices

$$\{(a,i) \mid a \in A,\ 0 < i < |\varphi(a)|\} \cup \{\omega\}$$

and edges

$$(a,i) \xrightarrow{b} (a,i+1) \text{ if } \varphi(a)_{i+1} = b$$

plus

- the edges $\omega \xrightarrow{b} (a,1)$ if b is the first letter of $\varphi(a)$,
- the edges $(a,|\varphi(a)|-1) \xrightarrow{b} \omega$ if b is the last letter of $\varphi(b)$
- and the loops $\omega \xrightarrow{b} \omega$ if $\varphi(a) = b \in A$.

Strong unambiguity and circular morphisms

A morphism $\varphi : A^* \to B^*$ is *circular* if it is injective and if for every $u, v \in B^*$

$$uv, vu \in \varphi(A^*) \Rightarrow u, v \in \varphi(A^*).$$

The set $\varphi(A)$ is called a *circular code*. There is a close connexion between strong unambiguity and circular morphisms.

Theorem (Restivo, 1975)

Let $\varphi : A^* \to B^*$ be a morphism. The following conditions are equivalent.

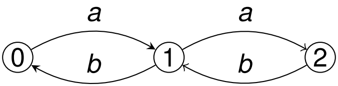
(i) *φ is circular.*

(ii) *The automaton $\mathscr{A}(\varphi)$ is strongly unambiguous.*

(iii) *The closure under the shift of $\varphi(A^{\mathbb{Z}})$ is of finite type.*

Relative unambiguity

Let X be a shift space on A. An automaton on A is *unambiguous on* X if, for every sequence $x \in X$, there is at most one path labeled x.



The above automaton is unambiguous on the *Thue-Morse shift*, which is the shift generated by the *Thue-Morse sequence* $x = \varphi^{\omega}(a)$ with $\varphi : a \mapsto ab, b \mapsto ba$.
Indeed, the words in $\mathscr{L}(X)$ of length 5 all contain aa or bb and there is only one path labeled aa or bb.

Mosse's Theorem

Let X be a shift space on A. A morphism $\varphi : A^* \to B^*$ is said to be *recognizable on X* if the automaton $\mathscr{A}(\varphi)$ is unambiguous on X.
If $\varphi : A^* \to A^*$ is a morphism, we denote by $X(\varphi)$ the set of sequences having all their blocks factors of some $\varphi^n(a)$ for some $n \geq 0$ and $a \in A$.

Theorem (Mosse, 1992)

A primitive aperiodic morphism φ is recognizable on the shift $X(\varphi)$.

Examples

- The automaton associated with the Fibonacci morphism is the golden mean automaton. It is strongly unambiguous. Thus the Fibonacci morphism (and actually every circular morphism) is recognizable on the full shift.
- The Thue-Morse morphism is recognizable on the Thue-Morse shift. It is actually recognizable on every minimal shift X such that aa or bb is in $\mathscr{L}(X)$.

Generalizations of Mosse's Theorem

A first generalization was proved by Bezuglyi, Kwiatowski, Medinets (2009): every aperiodic morphism φ is recognizable on $X(\varphi)$.
An automaton is unambiguous on X *for aperiodic points* if for every aperiodic $x \in X$ there is at most one path labeled x. A morphism $\varphi : B^* \to A^*$ is recognizable on X for aperiodic points if the automaton $\mathscr{A}(\varphi)$ is unambiguous on X for aperiodic points.
The following statement generalizes Mosse's Theorem.

Theorem (Berthé, Steiner, Thuswaldner, Yasawi, 2019)

A morphism $\varphi : A^* \to A^*$ is recognizable on $X(\varphi)$ for aperiodic points.

Indecomposable morphisms

A morphism $\varphi : B^* \to A^*$ is *indecomposable* if for every $\alpha : C^* \to A^*$ and $\beta : B^* \to C^*$ such that $\varphi = \alpha \circ \beta$, one has $\mathrm{Card}(C) \geq \mathrm{Card}(B)$.
The following is also proved by Berthé et al. (2019). It also follows from a result of Karhumaki, Manuch (2002)

Theorem

An indecomposable morphism $\varphi : B^* \to A^*$ is unambiguous on $A^{\mathbb{Z}}$ for aperiodic points.

A new proof of Mosse's Theorem

A new (and simpler) proof of Mosse's Theorem (and also of its generalization by Berthé et al.) can be derived from the statement concerning indecomposable morphisms. It relies on two simple lemmas.

Lemma

Let $\sigma : A^* \to A^*$ be a morphism. Every point y in $X(\sigma)$ has a σ-representation $y = S^k(\sigma(x))$ with $x \in X(\sigma)$.

Lemma

Let $\sigma : A^* \to A^*$ be an aperiodic morphism. For every aperiodic point x in $X(\sigma)$, the point $\sigma(x)$ is aperiodic.

Every morphism has a power which can be decomposed as $\sigma = \alpha \circ \beta$ with $\tau = \beta \circ \alpha$ indecomposable. If $x \in X(\sigma)$ is aperiodic, then $\beta(x)$ is in $X(\tau)$ and is aperiodic. Thus it has only one τ-representation.

Coded shifts

As defined by Blanchard and Hansel (1986), a coded shift is a shift space X such that $\mathscr{L}(X)$ is the set of factors of C^* for some language C. We say that X is coded by C.

The golden mean shift and the even shift are coded shifts.

The Fibonacci shift is not a coded shift (because in a coded shift, the set of periodic points is dense).

Coded shifts are defined equivalently by countable strongly connected automata (when C is finite, we can take the automaton associated with a morphism $\varphi : B^* \to A^*$ such that $C = \varphi(B)$).

Proposition

A shift space X is a coded shift if and only if it is the closure of the set of labels of infinite paths in a countable strongly connected automaton.

Coded shifts are a natural generalization of irreducible sofic shifts, which are those recognized by finite graphs.

Reversible automata

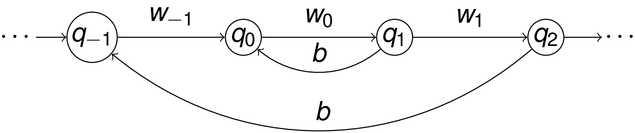
An automaton is called *reversible* if it is both deterministic and co-deterministic. The following result is from (Fiebig and Fiebig, 1992).

Theorem

Every coded shift is recognized by a countable strongly connected automaton which is reversible.

Example

Let X be the golden mean shift and let $(w_n)_{n\in\mathbb{Z}}$ be an enumeration of the words in $a\mathscr{L}(X)a$. A reversible automaton is represented below.



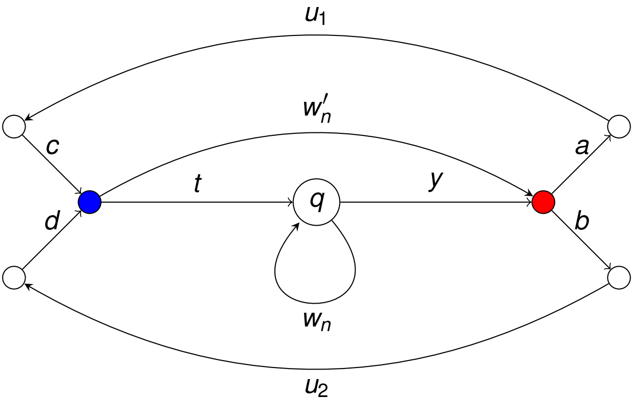
No finite reversible automaton recognizes the golden mean shift. Indeed, by a result of Pin (1992), a language L can be recognized by a finite reversible automaton if and only if for every $u, v, w \in A^*$, one has

$$uv^+w \subset L \Rightarrow uw \in L.$$

Thus the golden mean shift cannot be recognized by a finite reversible automaton.

Sketch of the proof

Assume that X is infinite and fix some state q of a countable strongly connected automaton recognizing X. Let $(w_n)_{n\in\mathbb{Z}}$ be an enumeration of the paths around q. Set $w'_n = tw_ny$.

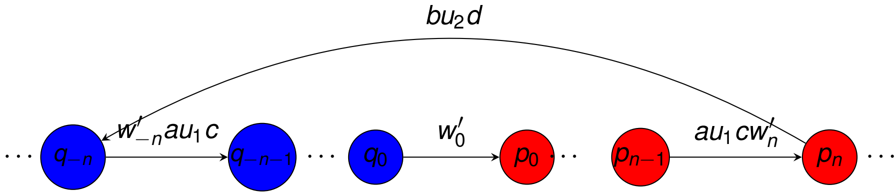


The new automaton

We build a new automaton starting from the skeleton

$$\cdots(w'_{-n}au_1c)\cdots(w'_{-1}au_1c)w'_0\cdot(au_1cw'_1)(au_1cw'_2)\cdots$$

and adding back edges to obtain strong connectedness.



Unambiguously coded shifts

A coded shift X defined by a language C is said to be *unambiguously coded* by C if for every $x \in X$ there exists at most one pair of a sequence $(c_n)_{n\in\mathbb{Z}}$ and an integer k with $0 \leq k < |c_0|$ such that

$$x = S^k(\cdots c_{-1} \cdot c_0 c_1 \cdots).$$

As an equivalent formulation, X is unambiguously coded by C if for every $x \in X$ there is at most one pair of a sequence $(c_n)_{n\in\mathbb{Z}}$ and a factorization $c_0 = ps$ with s nonempty such that

$$x = \cdots c_{-2} c_{-1} p \cdot s c_1 c_2 \cdots$$

This implies that C is a circular code. The converse is true when C is finite.

Example

The even shift is unambiguously coded. This is not true for $C = \{a, bb\}$ since the sequence $x = b^\infty$ has two factorizations. But it becomes true if we choose the prefix code $C' = (bb)^* a$.

Recall that an automaton is strongly unambiguous if the labelling of bi-infinite paths is injective, that is, it has at most one bi-infinite path with a given bi-infinite label.

Proposition

A shift space is unambiguously coded if and only if it can be recognized by a strongly connected and strongly unambiguous countable automaton.

The following result is due to Doris Fiebig (unpublished).

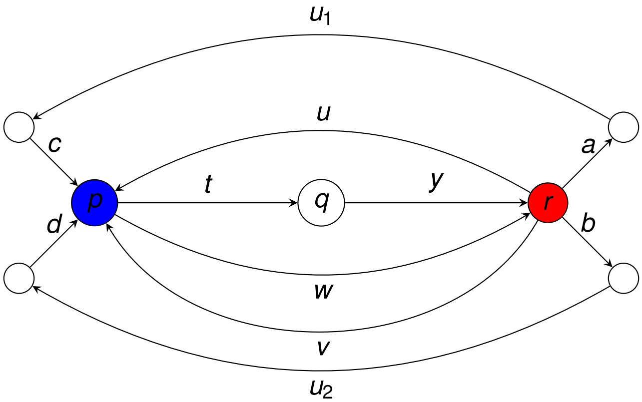
Theorem

Every coded shift is unambiguously coded.

The construction is effective and it implies that the topological entropy as well as the topological pressure of a coded shift (given by some computable code C) are computable (Burr et al. 2021).

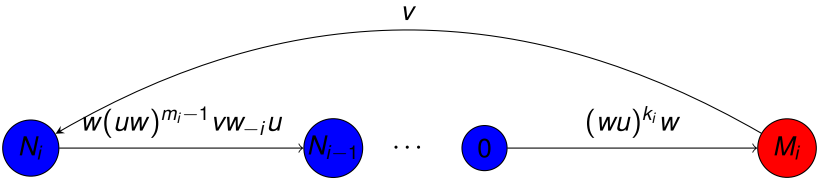
Idea of the proof

Start as in the previous proof. Let $(w_i')_{i\geq 1}$ be an enumeration of the labels of paths around q. Set $u = au_1c$, $v = bu_2d$, $w = ty$ and $w_{-i} = tw_i'y$.



The new automaton

For large enough integers $m_i \geq 0$ and $k_1 = m_1 < k_2 < \ldots$, it can be proved the automaton below is strongly unambiguous (the proof is not easy).



Synchronizing words

Given a prefix code C, a word $w \in C^*$ is *synchronizing* if for every $u, v \in A^*$, one has

$$uwv \in C^* \Rightarrow uw, v \in C^*. \tag{1}$$

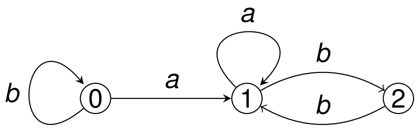
A prefix code C on the alphabet A is *synchronized* if there is a synchronizing word.
A shift space is said to be a *synchronized coded shift* it it can be defined by a synchronizing prefix code.

Synchronized automata

A word w is *synchronizing* for an automaton if there is a unique vertex reached by the paths labeled w. An automaton is *synchronized* if there is a synchronizing word. The following result is essentially due to Fischer (1975).

Proposition

An irreducible shift space X is a synchronized coded shift if and only if the minimal automaton of $\mathscr{L}(X)$ has a unique maximal strongly connected component which is synchronized.



It is well known that every irreducible sofic shift is a synchronized coded shift.

The following result is a consequence of the general result seen above. However, its proof builds a synchronizing automaton which remains finite for a sofic shift.

Theorem

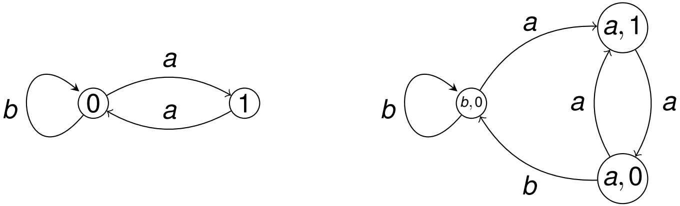
Every synchronized coded shift is unambiguously coded.

The proof goes as follows. Let X be a shift coded by a synchronized prefix code C. Let $w \in C^*$ be a synchronizing word of length n.

We consider the following automaton. The states are the pairs (u, p) formed of a word of length n in $\mathscr{L}(X)$ and an element p of the set P of states of the minimal automaton of C^*. Next, set
$(u, p) \cdot a = (v, p \cdot a)$ where v is such that $ua = bv$ for some letter b.
Since w is synchronizing, there is a state (w, q_w) in Q such that a path ends in (w, q_w) if and only if its label ends with w.
Let C' be the set of labels of simple paths from (w, q_w) to itself (such a path is simple if it does not pass by (w, q_w) in between). Then X is unambiguously coded by C'.

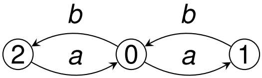
Example 1

The even shift is coded by $C = \{b, aa\}$. The letter b is synchronizing for C and the prefix code $C' = b \cup a(aa)^*ab = (aa)^*b$ is the result of the construction in the proof.

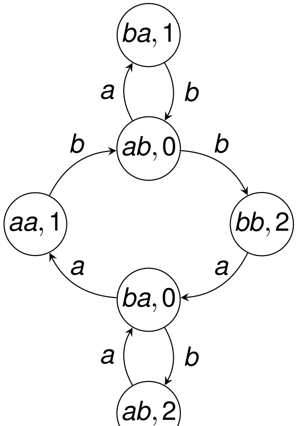


Example 2

Set $C = \{ab, ba\}$. The minimal automaton of C^* is represented below.

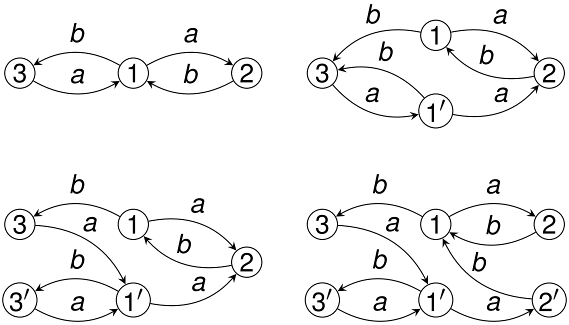


The automaton build in the proof is represented below.



State splitting

Another construction uses state-splitting (input split). On the example below, state 1 is first split into 1 and $1'$ (which have the same output and share the input of 1). Then 3 is split into 3 and $3'$ and finally 2 into 2 and $2'$.



The result is the same as by the other method.

Open problems

- Let $\varphi : A^* \to B^*$ be an indecomposable morphism. Is the number of periodic points which are the label of more than one path in the automaton $\mathscr{A}(\varphi)$ bounded by $\mathrm{Card}(A)$ (Karhumaki, Manuch, Plandowski, 2003)?
- Find a simpler proof of Fiebig's Theorem (every coded shift is unambiguously coded).
- Let $\varphi : A^* \to B^*$ be a morphism. If there is a sequence $x \in B^{\mathbb{Z}}$ with k disjoint factorizations, then $\varphi = \alpha \circ \beta$ with $A^* \xrightarrow{\beta} C^* \xrightarrow{\alpha} B^*$ and $\mathrm{Card}(C) \leq \mathrm{Card}(A) - k + 1$ (Karhumaki, Manuch, 2002)?

For more open problems, see Dimension Groups and dynamical systems (Fabien Durand and D.P.), Cambridge (2021), to appear.