# Master 2 Mathematics and Computer Science
# Symbolic Dynamics. Lecture 1


Marie-Pierre Béal

University Gustave Eiffel
Laboratoire d'informatique Gaspard-Monge UMR 8049

# Overview

- Shift spaces. Shifts of finite type. Sofic shifts.
- Irreducible sofic shifts. Minimal deterministic presentation.

# Words and languages

Let $A$ be a finite alphabet. The elements of $A$ are called letters or symbols.

A word on $A$ is a finite sequence of elements of $A$, denoted by $a_0 a_1 \cdots a_{n-1}$.

The set of words on $A$ is denoted by $A^*$, the *empty word* by $\varepsilon$, and he set of nonempty words by $A^+$.

A word $u$ *occurs in* a word $w$, or is a *block* of $w$ if $w = pus$ for some words $p, s$.

A *language* is a set of words.

# Sequences

Let $A$ be a finite alphabet.

A two-sided sequence is a sequence $x = (x_n)_{n \in \mathbb{Z}} \in A^{\mathbb{Z}}$.

For $i \leq j$, we write

$$x_{[i,j]} = x_i x_{i+1} \cdots x_j \text{ and } x_{[i,j)} = x_i x_{i+1} \cdots x_{j-1}.$$

A word $u$ *occurs in* a sequence $x$ if $u = x_{[i,j)}$ for some $i, j$. One also says that $u$ is a *block* of $x$, and that the integer $i$ is an *occurrence* of $u$ in $x$.

# Topology and shift transformation

The set $A^{\mathbb{Z}}$ of two-sided infinite sequences of elements of $A$ is a metric space for the distance defined for $x \neq y$ by $d(x,y) = 2^{-r(x,y)}$ where

$$r(x,y) = \inf\{|n| \mid n \in \mathbb{Z}, x_n \neq y_n\}. \tag{1}$$

The topology induced by this metric coincides with the product topology on $A^{\mathbb{Z}}$, using the discrete topology on $A$.
Since a product of compact spaces is compact, $A^{\mathbb{Z}}$ is a compact metric space.

Let $S$ denote the *shift transformation*, defined for $x \in A^{\mathbb{Z}}$ by $S(x) = y$ if $y_n = x_{n+1}$ for $n \in \mathbb{Z}$. It is continuous and one-to-one from $A^{\mathbb{Z}}$ to itself.

# Shift spaces

A set $X \subseteq A^{\mathbb{Z}}$ is *shift-invariant* if $S(X) = X$ (or, equivalently $S^{-1}(X) = X$).

A *shift space* on the alphabet $A$ is a subset of $A^{\mathbb{Z}}$ which is

(i) topologically closed,

(ii) shift-invariant.

For a language $F \subseteq A^*$, the set of sequences $x \in A^{\mathbb{Z}}$ such that no word of $F$ occurs in it is denoted by $\mathsf{X}_F$.

**Proposition**

*A set $X \subseteq A^{\mathbb{Z}}$ is a shift space if and only if $X = \mathsf{X}_F$ for some $F \subseteq A^*$.*

**Proof.**

Exercise. □

# Solution

Assume first that $X$ is a shift space on $A$ and let $F$ be the set of words on $A$ that do not occur in the elements of $X$. Then $X \subseteq \mathsf{X}_F$. Conversely, let $x \in \mathsf{X}_F$. For every $n \geq 1$, the word $x_{[-n,n]}$ is not in $F$ and is thus a block of some $y^{(n)} \in X$. Since $X$ is shift-invariant, we may choose $y^{(n)}$ such that $y^{(n)}_{[-n,n]} = x_{[-n,n]}$. The sequence $y^{(n)}$ converges to $x$, and since $X$ is closed, this implies $x \in X$.

Conversely, consider $X = \mathsf{X}_F$. Then $X$ is clearly closed and shift-invariant, and thus it is a shift space.

# Examples

The *full shift* $A^{\mathbb{Z}}$.

The *golden mean shift* is the set $X$ of two-sided infinite sequences on $A = \{a, b\}$ with no consecutive $b$.

In other terms, $X = \mathsf{X}_F$ with $F = \{bb\}$.

A *shift of finite type* is a shift $X = \mathrm{X}_F$ for some finite set $F$.
The golden mean shift is a shift of finite type.

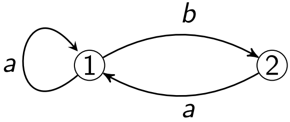
An easy way to define a shift space is to use a finite graph $\mathcal{A}$ with edges labeled by letters in $A$. Then the set $\mathrm{X}(\mathcal{A})$ of labels of two-sided infinite paths in $\mathcal{A}$ is easily seen to be shift-invariant.

It can be shown to be also closed (Exercise).

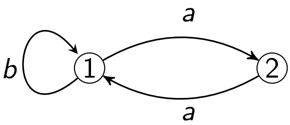
Thus, it is a shift space, called a *sofic* shift.

A shift of finite type is sofic but the converse is not true (Exercise).

# Example of a shift of finite type: the golden mean shift



# Example of a sofic shift: the even shift



# $X(\mathcal{A})$ is a shift space: Solution

Let $\mathcal{A}$ be a graph with edges labeled by letters in $A$.

Let $E$ be the set of edges of the graph $\mathcal{A}$. The set of bi-infinite paths in $\mathcal{A}$ is the shift $\mathsf{X}_F$ on the alphabet $E$, with $F$ being the words of length 2 of the form $ef$, where $e, f$ are not consecutive. Thus, it is a shift space.

The set $X(\mathcal{A})$ is the image of $\mathsf{X}_F$ under the map which assigns to a path its label. Since this map is continuous, the conclusion follows.

# A shift of finite type is sofic: solution

Let $X = \mathsf{X}_F$ with $F \subseteq A^*$ a finite set. Let $\mathcal{A} = (Q, I, T)$ be a trim automaton recognizing the language $A^* \setminus A^*FA^*$. The set of labels of two-sided infinite paths in $\mathcal{A}$ is equal to $X$.
The automaton is *trim* if every state is accessible and coaccessible.
Note: the proof holds when $F$ is only regular.

# A shift of finite type is sofic: another solution

Let $X = \mathsf{X}_F$ with $F \subseteq A^*$ a finite set.
Let $n$ be the maximal size of words in $F$.
Let $\mathcal{A}$ be the graph whose states are the words of length $n-1$ that do not contain any word of $F$, and with edges

$$a_0 a_1 \dots a_{n-2} \xrightarrow{a_0} a_1 \dots a_{n-2} a,$$

where $a_0 a_1 \dots a_{n-2} a$ does not contain any word of $F$. The set of labels of two-sided infinite paths in $\mathcal{A}$ is equal to $X = \mathsf{X}_F$.

Example with $F = \{bb\}$ on the board.

# The even shift is not of finite type: solution

Assume that $X = \mathsf{X}_F$ with $F \subseteq A^*$ formed of words of length at most $n$.
Then every block of length $n$ of $x = {}^{\omega}a \cdot ba^{2n+1}ba^{\omega}$ is in $\mathcal{B}_n(X)$, and thus $x$ has no block in $F$, a contradiction since $x \notin X$.

# Language of a shift space

If $X$ is a shift space, the set of blocks of sequences in $X$ is denoted by $\mathcal{B}(X)$. The set of blocks of length $n$ of sequences in $X$ is denoted by $\mathcal{B}_n(X)$.
A language $L$ is called *factorial* if it contains the words occurring as blocks in its elements, that is, if $uvw \in L$, then $v \in L$.
It is *extendable* if every $u \in L$ is *extendable*, that is, there are letters $a, b \in A$ such that $aub \in L$.

**Proposition**

*The language of a shift space is factorial and extendable. Conversely, for every factorial and extendable language $L$, there is a unique shift space $X$ such that $\mathcal{B}(X) = L$. It is the set $\mathrm{X}(L)$ of sequences $x \in A^{\mathbb{Z}}$ with all their blocks in $L$. For every factorial and extendable language $L$ and every shift space $X$, the following equalities hold: $\mathcal{B}(\mathrm{X}(L)) = L$, and $\mathrm{X}(\mathcal{B}(X)) = X$.*

# Solution

Let $L$ be a factorial extendable language. Let $\mathsf{X}(L)$ be the set of all sequences with all their blocks in $L$. Clearly, $\mathsf{X}(L)$ is a shift space and $\mathcal{B}(\mathsf{X}(L)) \subseteq L$.
For $u \in L$, since $L$ is extendable, there are sequences $(a_n)_{n\geq 0}$ and $(b_n)_{n\geq 0}$ such that $a_n \cdots a_0 u b_0 \cdots b_n \in L$. Since $L$ is factorial, the sequence $\cdots a_0 \cdot u b_0 b_1 \cdots$ is in $X(L)$. This shows that $L \subseteq \mathcal{B}(\mathsf{X}(L))$, and thus that $\mathcal{B}(\mathsf{X}(L)) = L$.

The second equality holds because for any $x \in \mathsf{X}(\mathcal{B}(X))$, one has $x_{[-n,n]} \in \mathcal{B}(X)$ and thus $x$ belongs to the closure of $X$, whence to $X$. Thus, the map $L \mapsto \mathsf{X}(L)$ is the inverse of the map $X \mapsto \mathcal{B}(X)$.

# Example

Find an example of a language $L$ such that $\mathcal{B}(\mathsf{X}(L)) \neq L$.

# Solution

Set $L = a^*ba^*$. Then $L$ is not factorial and $X(L)$ is empty.

Let $X$ be a shift space. For two words $u, v$ such that $uv \in \mathcal{B}(X)$, the set

$$[u \cdot v]_X = \{x \in X \mid x_{[-|u|,|v|)} = uv\}$$

is nonempty. It is called the *cylinder* with basis $(u, v)$. For $v \in \mathcal{B}(X)$, we also define

$$[v]_X = \{x \in X \mid x_{[0,|v|)} = v\}$$

in such a way that $[v]_X = [\varepsilon \cdot v]_X$. The set $[v]_X$ is called the *right cylinder* with basis $v$.

The open sets contained in $X$ are the unions of cylinders and the clopen sets are the finite unions of cylinders (Exercises).

# Solutions

Every cylinder is an open set, and thus every union of cylinders is open. Next, in any metric space, every open set is a union of open balls. But the open balls in $X$ are cylinders.

Every cylinder is a clopen set because the complement of $[u \cdot v]$ is the union of the cylinders $[u' \cdot v']$ with $u' \neq u$ of the same length as $u$ and $v = v'$, or $u' = u$ and $v' \neq v$ of the same length as $v$. Conversely, if $U$ is clopen, it is, as an open set, a union of cylinders $[u \cdot v]$ for a set of pairs $(u, v)$ such that $uv \in \mathcal{B}(X)$. Since $U$ is closed, it is compact. Thus, there is a finite set of pairs $(u, v)$ such that the union is equal to $U$.

# Irreducible shift

A nonempty shift space $X$ is *irreducible* if, for every $u, v \in \mathcal{B}(X)$, there is a word $w$ such that $uwv \in \mathcal{B}(X)$.

## Example

The golden mean shift $X$ is irreducible. Indeed, if $u, v \in \mathcal{B}(X)$, then $uav \in \mathcal{B}(X)$.

## Example

The shift $X$ formed of the labels of two-sided infinite paths in the graph below is reducible. Indeed, there is no word $u$ such that $bua \in \mathcal{B}(X)$.

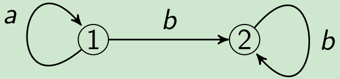



Figure: A reducible shift.

# Uniformly recurrent shift

A nonempty shift space $X$ is *uniformly recurrent* if for every $w \in \mathcal{B}(X)$ there is an integer $n \geq 1$ such that $w$ occurs in every word of $\mathcal{B}_n(X)$.

As an equivalent definition, a shift space $X$ is uniformly recurrent if for every $n \geq 1$ there is an integer $N = R_X(n)$ such that every word of $\mathcal{B}_n(X)$ occurs in every word of $\mathcal{B}_N(X)$. The function $R_X$ is called the *recurrence function* of $X$.

## Example

The golden mean shift $X$ is not uniformly recurrent since $b$ is in $\mathcal{B}(X)$ although $b$ does not occur in any $a^n \in \mathcal{B}(X)$.

# Deterministic automaton in symbolic dynamics

An automaton $\mathcal{A} = (Q, E)$ is a finite directed (multi)graph with edges labeled on $A$. The set of edges is included in $Q \times A \times Q$.

It is *trim* if each state has at least one outgoing edge and at least one incoming edge.

It is (uncomplete) *deterministic* if, for each state $p \in Q$ and each letter $a \in A$, there is at most one edge labeled by $a$ going out of $p$.

It is *irreducible* if its graph is strongly connected.

It is a *presentation* of a sofic shift $X$ if $X$ is the set of labels of bi-infinite paths of $\mathcal{A}$.

# Minimal automaton in symbolic dynamics

**Proposition**

*Every sofic shift has a trim deterministic presentation.*

**Proposition**

*Every irreducible sofic shift has a unique minimal deterministic presentation (irreducible deterministic and with the fewest number of states among these presentations).*

# Deterministic automaton in symbolic dynamics

**Proposition**

*Every sofic shift has a trim deterministic presentation.*

**Proof.**

Start with a trim presentation $\mathcal{A} = (Q, E)$. Apply the subset construction: $\mathcal{D} = (\mathfrak{P}(Q) \setminus \emptyset, \Delta)$. There is an edge $P \xrightarrow{a} P'$ in $\Delta$ is $P' = \{q \in Q \mid \exists p \in P, p \xrightarrow{a} q\}$. Start from $P = Q$. □

# Minimal automaton in symbolic dynamics

**Proposition**

*Every irreducible sofic shift has a unique minimal deterministic presentation.*

**Proof.**

Start with a trim deterministic presentation $\mathcal{A} = (Q, E)$ of $X$.
For $q \in Q$, let
$\mathrm{Fut}(q) = \{w \in A^* \mid \text{there is a path labeled by } w \text{ starting at } q\}$.
We say that $\mathcal{A}$ is *reduced* if $p \neq q$ implies $\mathrm{Fut}(p) \neq \mathrm{Fut}(q)$.
If $\mathcal{A}$ is not reduced, there are two states $p \neq q$ with $\mathrm{Fut}(p) = \mathrm{Fut}(q)$.
Merge $p, q$: let $\mathcal{A}' = (Q', E')$ with $Q' = Q \setminus \{p, q\} \cap \{(p, q)\}$, where $(p, q)$ is a new state and $E'$ is the set of edges $(\pi(r), a, \pi(s))$ with $(r, a, s)$ in $E$, $\pi(t) = t$ if $t \neq p, q$, $\pi(p) = \pi(q) = (p, q)$.
The automaton $\mathcal{A}'$ is still a trim deterministic presentation of $X$. □

**proof continued.**

Let $\mathcal{A}$ be a reduced trim deterministic presentation of $X$.
The automaton $\mathcal{A}$ has a *synchronizing word*: a word $w$ such that all paths labeled by $w$ end in the same state $q_w$.
Indeed, let
$\text{rank}(w) = \text{Card}\{q \mid \exists$ a path labeled by $w$ ending in $q\}$.
Let $w$ be a word of minimal non-null rank. Let us show that $\text{rank}(w) = 1$.
If $\text{rank}(w) > 1$, let $(p, w, q)$, $(p', w, q')$ be two paths with $q \neq q'$.
Since $\text{Fut}(q) \neq \text{Fut}(q')$, let $u$ be a word such that $u \in \text{Fut}(q) \setminus \text{Fut}(q')$ (or the converse).
Then, $\text{rank}(wu) < \text{rank}(w)$ and $\text{rank}(wu) \neq 0$.

□

**proof continued.**

Let $\mathcal{B}$ be the irreducible part of $\mathcal{A}$ containing $q_w$. Then $\mathcal{B}$ is an irreducible deterministic presentation of $X$ that is reduced.
Indeed, let $u$ be a block in $\mathcal{B}(X)$. There are blocks $v, v'$ such that $wvuv'w \in \mathcal{B}(X)$. Hence, there is a path in $\mathcal{A}$ labeled $vuv'w$ from $q_w$ to $q_w$, implying that $u$ is the label of a path in $\mathcal{B}$. Conversely, labels of bi-infinite paths of $\mathcal{B}$ belong to $X$.

Finally, let $\mathcal{C} = (Q', E')$ be another reduced irreducible deterministic presentation of $X$.
Let $w$ be a synchronizing word for $\mathcal{B}$ and $w'$ be a synchronizing word for $\mathcal{C}$. There is a word $u$ such that $wuw' \in \mathcal{B}(X)$. Thus $z = wuw'$ is a synchronizing word for both $\mathcal{B}$ and $\mathcal{C}$.
We define a bijection $f$ from $Q$ to $Q'$ as follows: $f(q_z) = q'_z$. If $q \in Q$ and $u$ is the label of a path from $q_z$ to $q$, we define $f(q)$ as the end of the unique path labeled by $u$ going out of $q'_z$ in $\mathcal{C}$.

□

# Local automaton

A deterministic automaton $\mathcal{A} = (Q, E)$ is *local* if there is an integer $n$ such that, for each word $w$ of length $n$, all paths labeled by $w$ end in the same state $q_w$.

**Proposition**

*An irreducible shift $X$ is of finite type if and only if its minimal deterministic automaton is local.*

**Proof.**

Exercise. □

# solution

**Proof.**

Let $X = \mathsf{X}_F$ be an irreducible shift of finite type. Without loss of generality, we may assume that all words of $F$ have the same length $n$.
Let $\mathcal{A} = (Q, E)$, where $Q$ is the set of words of length $n-1$ with edges

$$a_0 a_1 \dots a_{n-2} \xrightarrow{a} a_1 \dots a_{n-2} a,$$

where $a_0 a_1 \dots a_{n-2} a \notin F$. We keep only the trim part of this automaton.

Then $\mathcal{A}$ is deterministic and irreducible. Indeed, let $p = u$, $q = v$. Then $u$, $v$ are blocks of $X$ (say why). Since $X$ is irreducible, there is a word $w$ such that $uwv \in \mathcal{B}(X)$. By construction, there is a path from $p$ to $q$ in $\mathcal{A}$ labeled by $wv$. □

**Proof.**

The automaton $\mathcal{A}$ is local. Indeed, by construction, any path labeled by $w$ of length $n-1$ ends in the state $w$.
Since $\mathcal{A}$ is local, after a reduction (two states with the same future are identified), it remains local.
The (unique) minimal deterministic automaton of $X$ can be obtained with a reduction of $\mathcal{A}$. It is thus local.

Conversely, if the minimal deterministic automaton $\mathcal{B}$ of $X$ is local: there is an integer $k \geq 0$ such that for each $w$ of length $k$, all paths of $\mathcal{B}$ labeled by $w$ end in the same state $q_w$.
Let $F$ be the set of words of length $k+1$ that do not label any path in $\mathcal{B}$.
Then $X = \mathsf{X}_F$ (say why).

□

# Local automaton

**Proposition**

*An irreducible deterministic automaton is local if and only if it has at most one cycle with a given label.*

**Proof.**

Exercise. □

cycle : path $p = p_0 \xrightarrow{a_0} p_1 \xrightarrow{a_1} p_2 \ldots \xrightarrow{a_{m-1}} p_m = p$.
$m$ is the length of the cycle.

# Solution

**Proof.**

Let $\mathcal{A}$ be a deterministic irreducible automaton. If $\mathcal{A}$ has two cycles sharing the same label $w$.

$$p = p_0 \xrightarrow{a_0} p_1 \xrightarrow{a_1} p_2 \ldots \xrightarrow{a_{m-1}} p_m = p',$$
$$p' = p'_0 \xrightarrow{a_0} p'_1 \xrightarrow{a_1} p'_2 \ldots \xrightarrow{a_{m-1}} p'_m = p'.$$

We have $p_i \neq p'_i$ for some $0 \leq i < m$.
Since $\mathcal{A}$ is deterministic, $p_i \neq p'_i$ for all $0 \leq i < m$.
Then, for any $n = k|w| + j$, $0 \leq j < |w|$, $w^k w_{[0,j)}$ is the label of a path ending in $p_j$ and of a path ending in $p'_j \neq p_j$.
Thus, $\mathcal{A}$ cannot be local.

□

# Solution

**Proof.**

Conversely, if $\mathcal{A}$ is not local, then, for any integer $n$ there are two paths labeled by a word $w^{(n)}$ of length $n$ ending in distinct states. We choose $n = (\text{Card } Q)^2$. These two paths are

$$p = p_0 \xrightarrow{a_0} p_1 \xrightarrow{a_1} p_2 \ldots \xrightarrow{a_{n-1}} p_n = p',$$
$$p' = p'_0 \xrightarrow{a_0} p'_1 \xrightarrow{a_1} p'_2 \ldots \xrightarrow{a_{n-1}} p'_n = p'.$$

If $p_i = p'_i$ for some $0 \leq i < n$, then $p = p'$, a contradiction. Hence $p_i \neq p'_i$ for all $0 \leq i < n$. By the pigeonhole principle, there are $0 \leq i < j < n$ such that $(p_i, p'_i) = (p_j, p'_j)$, implying the existence of two cycles sharing the same label.

□