Words and Automata, Lecture 1
Symbolic dynamics

Dominique Perrin

21 novembre 2017

Outline of the course

- Symbolic dynamical systems
- Substitution shifts
- Dimension groups
- Ordered cohomology

Objectives of the course

- General : Give an introduction to symbolic dynamics, combinatorics on words and automata. All aspects are not covered but the travel gives a general idea of the topics.
- Particular : Focus on the notion of dimension group as a powerful invariant of minimal subshifts (invariant means invariant under isomorphism).
- Practical : Discover the various tools to compute dimension groups (and other things also) : Rauzy graphs, return words, higher block presentations,...

Outline of this lecture

- Symbolic dynamical systems
- Recurrent and uniformly recurrent systems
- Sturmian systems
- Return words

Let A be a finite set called the alphabet. We denote by A^* the set of finite words on A and by $A^{\mathbb{Z}}$ the corresponding set of two-sided infinite words. The set $A^{\mathbb{Z}}$ is a metric space for the distance $d(x, y) = 2^{-r(x,y)}$ for $r(x, y) = \max\{n > 0 \mid x_i = y_i \text{ for } -n < i < n\}$ (with $r(x, y) = \infty$ if $x = y$).
The shift transformation is defined for $x = (x_n)_{n\in\mathbb{Z}}$ by $y = Tx$ if

$$y_n = x_{n+1}$$

for $n \in \mathbb{Z}$. It is a continuous map from $A^{\mathbb{Z}}$ onto $A^{\mathbb{Z}}$.

Symbolic dynamical systems

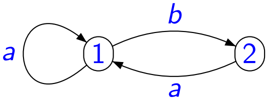
A set $X \subset A^{\mathbb{Z}}$ is closed if for any sequence $x^{(n)}$ in X converging to $x \in A^{\mathbb{Z}}$, one has $x \in X$.
A set $X \subset A^{\mathbb{Z}}$ is invariant by the shift if $T(X) = X$.
A symbolic dynamical system (also called a subshift or a shift space) on the alphabet A is a subset of $A^{\mathbb{Z}}$ which is

- closed
- invariant by the shift.

As an equivalent definition, given a set S of finite words (the forbidden blocks) a symbolic dynamical system is defined as the set X_S of two-sided infinite words which do not have a factor in S.

Example

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



The Fibonacci shift

Let $\varphi : A^* \to A^*$ be the substitution $a \mapsto ab, \; b \mapsto a$. Since $\varphi(a)$ begins with a, every $\varphi^n(a)$ begins with $\varphi^{n-1}(a)$.
The Fibonacci word is the right infinite word with prefixes all $\varphi^n(a)$.
The Fibonacci shift is the set of biinfinite words whose blocks are blocks of the Fibonacci word.
Forbidden blocks : $bb, aaa, babab, \cdots$.

The language of a subshift

Let (X, T) with $X \subset A^{\mathbb{Z}}$ be a subshift. Let $L(X) \subset A^*$ be the set of finite words which are factors (or blocks) of the elements of X (sometimes called the language of X). We denote by $L_n(X)$ the set words of length n in $L(X)$.

Shifts of finite type

A shift of finite type is a subshift defined by a finite set of forbidden blocks. Thus (X, T) is of finite type if the exists a finite set $S \subset A^*$ such that $L(X) = A^* \setminus A^*SA^*$.

The golden mean subshift is a subshift of finite type. It corresponds to the set of forbidden blocks $S = \{bb\}$.

Sofic shifts

A sofic shift on the alphabet A is the set of labels of two-sided infinite paths in a finite graph labeled by A.

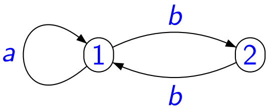
Proposition

Any shift of finite type is sofic.

Indeed, assume that (X, T) is defined by a finite set of forbidden blocks S. We may assume the S is formed of words all of the same length n. Let Q be the set of words of length n which are not in S. Let G be the graph on the set Q of vertices with an edge (u, v) labeled b if $u = aw$ and $v = wb$ for $a, b \in A$. Then X is the set of labels of two sided infinite paths in G.

Example

The even shift is the set of two-sided infinite paths in the graph below.



The set of forbidden blocks is the set of words ab^na with n odd.
The even shift is not of finite type.

Recurrent shifts

A subshift is recurrent if and only if for evey $u, v \in L(X)$ there is a w such that $uwv \in L(X)$.

Proposition

A sofic shift is recurrent if and only if it can be defined by a strongly connected graph.

The condition is clearly sufficient. The proof of its necessity uses additional knowledge (such as the minimal automaton of a sofic shift).

Minimal shifts

A subshift (X, T) is minimal (or uniformly recurrent)if and only if for every $u \in L(X)$ there is an $n \geq 1$ such that u is a factor of every word in $L_n(X)$.

Proposition

A subshift is minimal if and only if it does not contains properly any nonempy subshift.

Necessity : assume that $Y \subset X$ with X minimal. Let $u \in L(X)$. Then there is $n \geq 1$ such that u is a factor of any word in $L_n(X)$ and thus of any word in $L_n(Y)$. Thus $u \in L(Y)$. This shows that $L(X) = L(Y)$ and thus that $X = Y$.
Sufficiency : Consider a word $u \in L(X)$ such that for every $n \geq 1$ there is a word $w \in L_n(X)$ which has no factor equal to u. Use König's lemma to build a word $x \in X$ without factor u. Finally define Y to be the set of $x \in X$ without factor u.

Cylinders

For $w \in L(X)$, the set $[w] = \{x \in X \mid x_{[0,n-1]} = w\}$ is nonempty. It is called the cylinder with basis w. The clopen sets in X are the finite unions of cylinders.

Factor complexity

The factor complexity of the subshift (X, T) is the sequence

$$p_n(X) = \mathrm{Card}(L(X) \cap A^n).$$

The factor complexity of the golden mean shift is the Fibonacci sequence $1, 2, 3, 5, \ldots$ (arguing on the two kinds of factors, according to the last letter).
The factor complexity of the Fibonacci shift is the sequence $1, 2, 3, 4, \ldots$ (see below).

Theorem (Morse,Hedlund)

If $p_n(X) \leq n$ for some n, then X is finite.

Proof : If $p_1 = 1$, then $\mathrm{Card}(X) = 1$. Otherwise, consider an n such that $p_n = p_{n+1}$ (exists because p_n is nondecreasing and $p_1 \geq 2$). Then each factor of length n has a unique extension and thus X is finite.

Left and right special words

Let (X, T) be a subshift with $X \subset A^{\mathbb{Z}}$. For $w \in L(X)$, there is at least one letter $a \in A$ such that $wa \in L(X)$ and symmetrically, at least one letter $a \in A$ such that $aw \in L(X)$. The word w is called right-special if there is more than one letter $a \in A$ such that $wa \in L(X)$. Symmetrically, w is left-special if there is more than one letter $a \in A$ such that $aw \in L(X)$.

Examples

The right-special words for the golden mean shift are those ending with a.
The left special words for the Fibonacci shift are the prefixes of the Fibonacci word (reasoning by induction on its antecedent by the Fibonacci morphism).

Sturmian shifts

A recurrent subshift (X, T) on a binary alphabet is called Sturmian if it has complexity $p_n = n + 1$.
Equivalent definition : there is a unique right special word of each length.
Example : the Fibonacci shift is Sturmian.

Proposition

Any Sturmian subshift is minimal.

Consequence of the Morse, Hedlund Theorem.

One-sided symbolic dynamical systems

As a variant, we may consider the set $A^{\mathbb{N}}$ of one-sided infinite sequences with the one-sided shift defined by $y = Tx$ if $y_n = x_{n+1}$ for $n \geq 0$. Note that the one-sided shift is not one-to-one. Indeed, there are $\mathrm{Card}(A)$ one-sided sequences x such that $y = Tx$, differing by their first coordinate.

A one-sided symbolic dynamical system (or one-sided subshift) is a closed invariant subset of $A^{\mathbb{N}}$.

Let (X, T) be a (two-sided) subshift. Let $\theta : A^{\mathbb{Z}} \to A^{\mathbb{N}}$ be the natural projection. It induces a factor map from (X, T) onto the one-sided subshift (Y, S) where $Y = \theta(X)$. The one sided subshift (Y, S) is called the one-sided subshift associated to (X, T).

Return words

For $w \in L(X)$ a right return word to w is a word u such that wu is in $L(X)$, has w as a proper suffix and has no factor w which is not a prefix or a suffix. Symmetrically, a left return word to w is a word u such that uw is in $L(X)$, has w as a proper prefix and has no other factor w.

We denote by $\mathcal{R}_X(w)$ (resp. $\mathcal{R}'_X(w)$) the set of right (resp. left) return words to w.

Clearly a recurrent subshift (X, T) is minimal if and only if $\mathcal{R}_X(w)$ is finite for every $w \in L(X)$.

Higher block shifts

Let (X, T) be a subshift on the alphabet A and let $k \geq 1$ be an integer. Let $f : A_k \to L_k(X)$ be a bijection from an alphabet A_k onto the set $L_k(X)$ of blocks of length k of X. The map $\gamma_k : X \to A_k^{\mathbb{Z}}$ defined for $x \in X$ by $y = \gamma_k(x)$ if for every $n \in \mathbb{Z}$

$$y_n = f(x_n \cdots x_{n+k-1})$$

is the kth higher block code on X. The image $X^{(k)} = \gamma_k(X)$ is called the higher block shift of X. The higher blok code is an isomorphism of dynamical sytems and the inverse of γ_k is given by the map $\pi : A_k \to A$ which assigns to every $b \in A_k$ the first letter of $f(b)$.
We sometimes, when no confusion arises, identify A_k and $L_k(X)$ and write simply $y_0 y_1 \cdots = (x_0 x_1 \cdots x_{k-1})(x_1 x_2 \cdots x_k) \cdots$.

Consider again the golden mean shift (X, T). We have $L_3(X) = \{aaa, aab, aba, baa, bab\}$. Set $f : x \mapsto aaa, y \mapsto aab, z \mapsto aba, t \mapsto baa, u \mapsto bab$. The third higher block shift $X^{(3)}$ of X is the set of two-sided infinite paths in the graph below.