Codes in symbolic systems
DP, Marie-Pierre Béal

- Introduction
- Symbolic systems
- Entropy
- zeta function
- Codes
- Kraft's inequality
- The equality case
- Factorization

Introduction

Subject: interplay symbolic dynamics ↔ combinatorics on words.

Other attempts:

- Sturmian words (Morse, Hedlund)
- Unavoidable sets (Hansel et al.)
- De Bruijn graphs (Eduardo Moreno)

The Lothaire trilogy

- Combinatorics on Words, Addison Wesley, 1984
- Algebraic Combinatorics on Words, Cambridge, 2003
- Applied Combinatorics on Words, Cambridge, 2005
- Dynamical ...?

Symbolic systems

Subshift on A : closed and shift invariant subset of $A^{\mathbb{Z}}$.
Full shift : $A^{\mathbb{Z}}$.
Subshift of finite type : defined by a finite set of forbidden blocks.

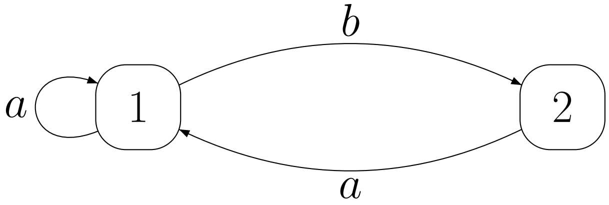

Figure 1: The golden mean system.

Sofic subshift : defined by a finite automaton (or a regular set of forbidden blocks).

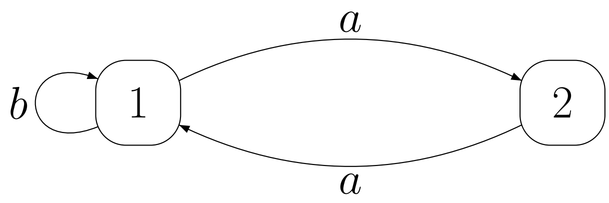

Figure 2: The even system.

Spectral constraints

The constraint $[d, k]$ for $d < k$: between two consecutive 1, at least k 0 and no more than d. It defines a subshift of finite type.

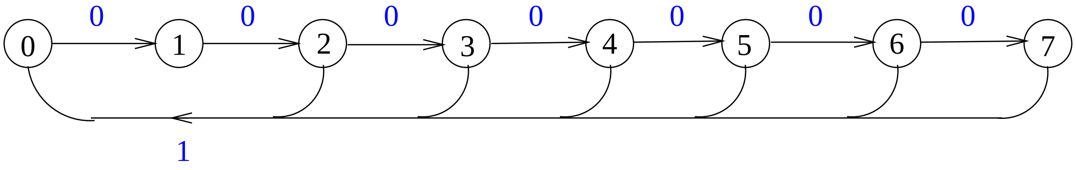

Figure 3: The $[2, 7]$ constraint

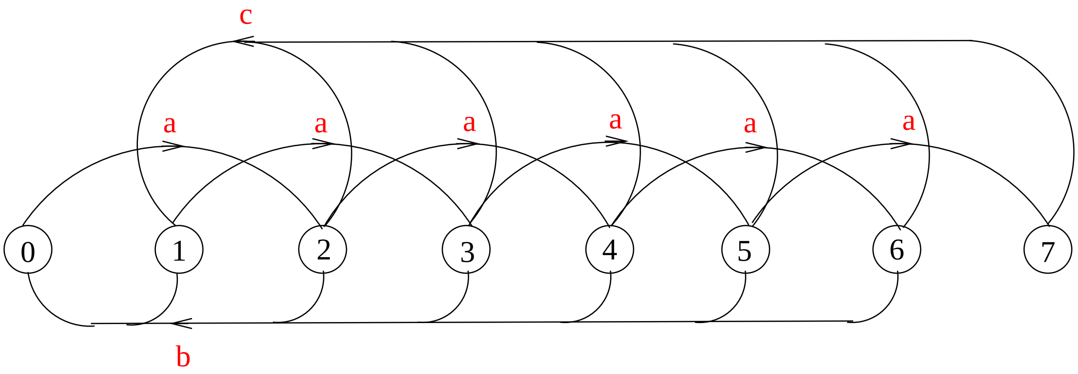

Figure 4: The $[2, 7]$ constraint on the alphabet $a = 00, b = 01, c = 10$

Entropy

The entropy of S is

$$h_S = \lim \frac{1}{n} \log u_n$$

whith u_n = number of admissible blocks of length n. Let ρ_S be such that

$$h_S = -\log \rho_S$$

For the full shift on k symbols, $\rho_S = 1/k$.
For the golden mean system, ρ_S is the inverse of the golden mean, i.e. such that $\rho + \rho_S^2 = 1$.

For the $[2,7]$ constraint, one has $h_S > 1/2$, i.e. $\rho_S < \sqrt{2}$. Thus the modified system obtained on blocks of length 2, satisfies $h_{S'} > 1$, i.e. $\rho_{S'} < 1/2$.

Zeta function

Let S be a subshift and let u_n be the number of points of period n :

$$u_n = \mathrm{card}\{x \in S \mid \sigma^n(x) = x\}.$$

Then

$$\zeta_S(z) = \exp \sum \frac{u_n}{n} z^n$$

is the zeta function of S. For a subshift of finite type realized as an edge shift on a graph G

$$\zeta(z) = \det(I - Mz)^{-1}$$

where M is the adjacency matrix of G (Manning, 1971). For the golden mean system, $\zeta(z) = \frac{1}{1-z-z^2}$.

Codes

Let S be a subshift on A. Let $F(S)$ denote the factors of S. A set $X \subset F(S)$ is

- a code if X^* is unambiguous.
- complete in S if $X \subset F(S) \subset F(X^*)$

caveat: complete $\not\Longrightarrow$ maximal: $X = \{ab\}$ is complete in $(ab)^\zeta$ although $X \subset \{ab, ba\}$ which is a code.

For example, $X = \{aa, ab, ba\}$ is a complete code for the golden mean system.

Note that we do not require $X^* \subset F(S)$ as in Restivo (1990). Our definition is more general and captures also the codes of paths of Reutenauer (1986).

Matrices

Let S be a sofic system recognized by a deterministic automaton on Q (the Shannon cover). We will work with the matrices

$$\mu(u)_{pq} = \begin{cases} u & \text{if } p \cdot u = q \\ 0 & \text{otherwise} \end{cases}$$

for $u \in A^*$ and $p, q \in Q$.

For the golden mean

$$\mu(a) = \begin{bmatrix} a & 0 \\ a & 0 \end{bmatrix}, \quad \mu(b) = \begin{bmatrix} 0 & b \\ 0 & 0 \end{bmatrix}$$

Series of matrices

Let π be the morphism obtained from μ through an assignment of positive real values to the elements of A.

For a set X, we denote

$$f_X(z) = \sum_{x \in X} \pi(x) z^{|x|}$$

For the golden mean

$$f_A(z) = \begin{bmatrix} pz & qz \\ pz & 0 \end{bmatrix}$$

If X is a code, then

$$f_{X^*}(z) = (I - f_X(z))^{-1}$$

The radius of convergence of $f_{X^*}(z)$ is the minimal root of $\det(I - f_X(z)) = 0$.

Admissible assigments

A generalization of Bernoulli distributions: an assignment π is admissible if $z = 1$ is the minimal root of $\det(I - f_A(z))$. As an example, the uniform assignment $\pi(a) = \rho_S$ for all $a \in A$ is admissible.

Let $A = \{a, b\}$ and let $\pi(a) = p, \pi(b) = q$.
For the full shift, π is admissible if $p + q = 1$.
For the golden mean, since

$$\det(I - f_A(z)) = \begin{vmatrix} 1 - pz & -qz \\ -pz & 1 \end{vmatrix} = 1 - pz - pqz^2,$$

π is admissible if $p + pq = 1$.

If π is admissible, $\rho(f_{A^*}(z)) = 1$.

If S is of finite type and π is uniform,

$$\zeta_S(z) = det(I - f_A(z/\rho_S))^{-1}.$$

Kraft's inequality

Theorem *Let S be an irreducible sofic shift. If $X \subset F(S)$ is a code, then* $\det(I - f_X(1)) \geq 0$ *for any admissible assignment π.*

For the code $X = \{ab, ba\}$ in the golden mean system, we have

$$f_X(1) = \begin{bmatrix} pq & pq \\ pq & 0 \end{bmatrix}$$

Since,

$$\det(I - f_X(1)) = \begin{vmatrix} 1 - pq & -pq \\ -pq & 1 \end{vmatrix} = 1 - pq - p^2q^2$$

the inequality reads

$$pq(1 + pq) \leq 1.$$

Let $1 - p_X(z) = \det(I - f_X(z/\rho_S))$.

Corollary *Let S be an irreducible sofic shift. If $X \subset F(S)$ is a code, then $p_X(\rho_S) \leq 1$.*

For the full shift on k symbols with the uniform assignment, we have $p_X(z) = \sum_{x \in X} z^{|x|}$ and thus the above is the classical Kraft inequality

$$\sum_{x \in X} k^{-|x|} \leq 1.$$

Proof

Since π is admissible,

$$\rho(f_{A^*}(z)) = 1.$$

Thus

$$\rho(f_{X^*}(z)) \geq 1.$$

Since X is a code,

$$f_{X^*}(z) = (I - f_X(z))^{-1}.$$

Thus $\det(I - f_X(z)) \neq 0$ for $0 \leq z \leq 1$. Hence $\det(I - f_X(z)) \geq 0$.

The Franaszek code

Allows to encode any binary sequence into one satisfying the $[2, 7]$ constraint at rate $1/2$. The code on the right is prefix with finite synchronization delay. The one on the left is a complete prefix code.

X	Y
ba	10
ca	11
aba	000
cba	010
aca	011
$acba$	0010
$aaca$	0011

The best way the see that the coded sequences satisfy the constraints is to remark that the codewords globally stabilize the set $\{2, 3\}$. Thus for the uniform assignment, $p_X(z) = p_Y(z) = 1 - 2z^2 - 3z^3 - 2z^4$. We have $p_X(1/2) = p_Y(1/2) = 1$ in accordance with the fact that $\rho_S < 1/2$

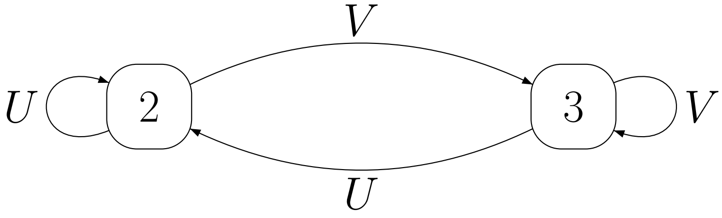

Figure 5: The action of the Franaszek code on the two poles 2 and 3 with $U = \{ba, aba, cba, acba\}$ and $V = \{ca, aca, aaca\}$.

The equality case

Theorem *Let S be an irreducible sofic shift, let π be an admissible assignment and let $X \subset F(S)$ be a regular code. The code X is S-complete if and only if* $\det(I - f_X(1)) = 0$.

For the code $X = \{aa, ab, ba\}$ in the golden mean system, the equality reads

$$1 - p^2 - 2pq + p^2q^2 = (1 + p - pq)(1 - p - pq) = 0.$$

Corollary *Let S be an irreducible sofic shift and let $X \subset F(S)$ be a regular code. The code X is complete in S if and only if $p_X(\rho_S) = 1$.*

Actually, for the above code, $p_X(z) = 1 - 3z^2 + z^4$.

Proof

Let X be complete in S. Since X is regular, each entry of $\mu(A^*)$ is contained in a finite union of two-sided residuals of entries of $\mu(X^*)$. This shows that $\rho(f_{X^*}(z)) = 1$, whence $\det(I - f_X(1)) = 0$.

Conversely, one has to show that if X is not complete, then $\rho(f_{X^*}(z)) > 1$ whence $\det(I-f_X(1)) > 0$.

Factorization

Let us denote by α the morphism obtained from μ by taking the commutative image of the elements. Then $p(X) = \det(I - \alpha(X))$ is, for any finite code X, a polynomial in $\mathbb{Z}[A]$.

The following solves a problem raised by Reutenauer, 1986, who proved the same result under an additional hypothesis.

Theorem *Let S be an irreducible sofic shift. When X is a finite S-complete code, the polynomial $p(X)$ is a multiple of the polynomial $p(A)$.*

For the golden mean, $p(A) = 1 - p - pq$ and for $X = \{aa, ab, ba\}$, we have

$$\begin{aligned} p(X) &= 1 - aa - 2ab + a^2b^2 \\ &= (1 + a - ab)(1 - a - ab). \end{aligned}$$

Actually, let $Y = A^2$. Then $1 - Y = (1 + A)(1 - A)$ whence the above since $\det(I + \alpha(A)) = 1 + a - ab$.

For $X = \{aa, ba, baa\}$, we have

$$\begin{aligned} p(X) &= 1 - a^2 - ab - a^2b \\ &= (1 + a)(1 - a - ab). \end{aligned}$$

Proof

The proof operates a first reduction to the case of an edge shift: the labels of all edges are distinct. In this case, $p(A)$ is irreducible in $\mathbb{Z}[A]$ (Reutenauer, 1986).

We select a letter $a \in A$ and write

$$p(A) = -aq + r$$

where q, r are polynomials in $\mathbb{Z}[A - a]$.

Then

$$p(X)q^n = p(A)s' + t'$$

for some $n \geq 0$ and some $s' \in \mathbb{Z}[A]$, $t' \in \mathbb{Z}[A - a]$.

There exists a ball B of radius $\epsilon > 0$ centered at $(\rho_S)_{b \in A-a}$ such that the assignment $\pi(b) = x_b$ for $x \in B$ with $\pi(a) = r(x)/q(x)$ is admissible.

Then $p(X) = p(A) = 0$ for any such assignment π and thus t' vanishes on B whence $t' = 0$.

Since $p(A)$ is irreducible, this forces $p(A) \mid p(X)$.

The noncommutative factorization theorem

Theorem (Reutenauer, 1984) *For any finite complete code $X \subset A^*$ there exists two polynomials $L, R \in \mathbb{Z}\langle A\rangle$ such that $X - 1 = L(A-1)R$.* Actually, for $Y = \{aa, ba, baa, bb, bba\}$, we have

$$1 - Y = (1 + a)(1 - A)(1 + b)$$

whence the above.