Synchronization delay

Dominique Perrin

March 14, 2007

Outline

Material prepared for the new edition of *Theory of Codes* (*Theory of Codes and Automata*, Jean Berstel, DP, Christophe Reutenauer)

- Synchronizing words
- Synchronization delay
- Local automata
- Completion
- Star-free closure
- Krieger's embedding theorem
- Nasu's masking lemma

Synchronizing words

A word $x \in X^*$ is synchronizing for $X \subset A^+$ if for all $u, v \in A^*$

$$uxv \in X^* \implies ux, xv \in X^*$$

Examples: The word a is synchronizing for the Fibonacci code

$$X = \{a, ba\}.$$

The word $x = abba$ is synchronizing for the Morse code

$$X = \{ab, ba\}$$

If x, y are synchronizing, then the pair (x, y) is synchronizing:

$$uxyv \in X^* \implies ux, yv \in X$$

Verbal synchronization delay

A code $X \subset A^*$ has verbal synchronization delay s if any word in X^s is synchronizing (Golomb and Gordon, 1965).
Examples:

- The Fibonacci code $X = \{a, ba\}$ has synchronization delay 1.
- The code $X = \{a, aba\}$ has synchronization delay 2.
- The Morse code $X = \{ab, ba\}$ does not have finite synchronization delay.

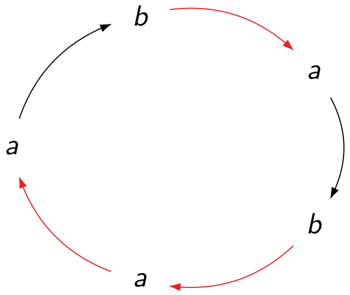
A comma free code is a set $X \subset A^n$ which has syncronization delay 1 (a word of X cannot be a nontrivial factor of a word of X^2).

Circular codes

A set X is a circular code if $uv, vu \in X^*$ imply $u, v \in X^*$.
Equivalently: any necklace has a unique decomposition in words of X.

Examples:

- The Fibonacci code $X = \{a, ba\}$ is circular.
- $X = b + ab^*c$ is circular (but not with finite synchronization delay).



Factorizations

A family $(X_i)_{i \in I}$ of subsets of A^* indexed by a totally ordered set I is a factorization of A^* if any word $w \in A^*$ can be written uniquely

$$w = x_1 x_2 \cdots x_n$$

with $x_i \in X_{j_i}$ and $j_1 \geq j_2 \geq \ldots \geq j_n$.
Example: the Lyndon factorization.

Theorem (Schützenberger, 1965)

If $A^* = \prod_{i \in I} X_i^*$ is a factorization, then each X_i is a circular code (and each conjugacy class meets exactly one X_i^*).

Limited codes

Let $p, q \geq 0$ be integers. A set $X \subset A^+$ is (p, q)-limited if for all $u_0, u_1, \ldots, u_{p+q}$ in A^*,

$$u_{i-1}u_i \in X^* \qquad (1 \leq i \leq p+q)$$

imply $u_p \in X^*$.
Example: X is $(1, 0)$-limited if $uv \in X^*$ implies $v \in X^*$, i.e. X^* is suffix-closed.
A limited code is circular.

Open question

Is every factor of a finite factorization $A^* = X_1^* X_2^* \ldots X_n^*$ limited?

yes for $n \leq 3$.

Local testability

A set $L \subset A^*$ is stricly locally testable (slt) if

$$L = T \cup (UA^* \cap A^*V) \setminus A^*WA^*$$

for finite sets T, U, V, W.

Theorem (Restivo, 1974)

For a finite code X the following conditions are equivalent.

(i) *X is circular.*

(ii) *X has finite synchronization delay.*

(iii) *X^* is strictly locally testable.*

In general (iii) $\Longrightarrow$ (ii) $\Longrightarrow$ (i).

Star-free sets

The star-free operations are the boolean operations and the product.

Theorem (Schützenberger, 1975)

If X is a code with finite synchronization delay, then X^* belongs to the star-free closure of X.

Proof:

$$X^* = 1 \cup X \cup \ldots \cup X^{s-1} \cup (X^s A^* \cap A^* X^s) \setminus W$$

where $W = \{w \in A^* \mid A^* w A^* \cap X^* = \emptyset\}$ has also the expression

$$W = (A^* \setminus A^* X^{2s+1} A^*) \cap (A^* \setminus F(X^{2s+2}))$$

Local automata

A finite automaton is local if there are integers s, t such that for any paths $p \xrightarrow{u} q \xrightarrow{v} r$ and $p' \xrightarrow{u} q' \xrightarrow{v} r'$, with $|u| = s$, $|v| = t$, one has $q = q'$.

Equivalent conditions for a strongly connected automaton $\mathcal{A}$:

(i) distinct cycles have distinct labels.

(ii) $\mathcal{A}$ is unambiguous and for long enough w, the relation

$$\varphi(w) = \{(p, q) \mid p \xrightarrow{w} q\}$$

has rank ≤ 1.

$X = \{aaca, aba, aca, , acba, ba, ca, cba\}$ is a circular prefix code.

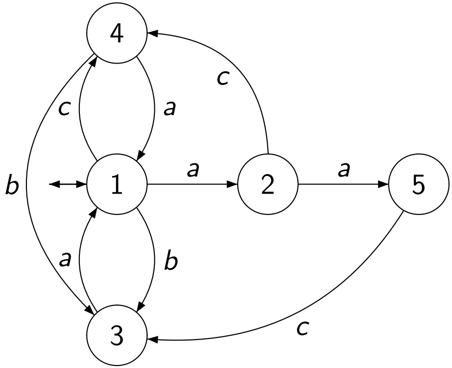


Figure: The minimal automaton of the Franaszek code.

The automaton is local with $s = 4$ and $t = 0$.

Literal synchronization delay

A word w is a constant for $L \subset A^*$ if the set of its contexts is a direct product:

$$\ell wr, \ell' wr' \in L \implies \ell' wr, \ell wr' \in L.$$

If $x \in X^*$ is a constant, then it is synchronizing. If (x, y) is a synchronizing pair, then xy is a constant.
A code $X \subset A^*$ has literal synchronization delay s if any word of A^s is a constant.
literal delay $\leq$ verbal delay $\leq 2Lmax$ (literal delay $+1$).

Theorem

The following conditions are equivalent for a code X.

(i) *X has finite literal synchronization delay.*

(ii) *X^* is strictly locally testable.*

(iii) *X^* is the stabilizer of a state in a local automaton.*

Proof

(i)$\Leftrightarrow$ (ii) take $U = X^*A^- \cap A^s$, $V = A^-X^* \cap A^s$ and $W = A^{s+1} \setminus F(X^*)$.
(i)$\Leftrightarrow$ (iii) For $\implies$, consider the minimal deterministic automaton $\mathcal{A} = (Q, i, T)$ of X. Then $\mathcal{A}^* = (Q \cup \omega, \omega, \omega)$ is local.
$\Leftarrow$ is clear.

Completion

Theorem (elaborated from Bruyère, 1998)

Any rational code with finite verbal (literal) deciphering delay is contained in a maximal one with the same delay.

Solution: the basis Y of the submonoid

$$M = (X^s A^* \cap A^* X^s) \cup X^*$$

Example: $X = \{a, ab\}$, $M = aA^* \cap A^*X$, $Y = (abb^+)^*X$.
For the literal delay:

$$M = (P_s A^* \cap A^* S_s) \cup X^*$$

Example: $X = \{a, ab\}$, $M = aA^*$, $Y = ab^*$.

Proof

In both cases, one has to prove that:

1. M is stable: $u, wv, uw, v \in M$ imply $w \in M$.

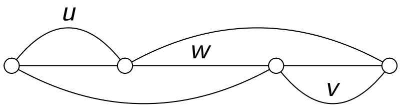


Figure: Proving that M is stable.

2. $X \subset Y$
3. Y is complete with synchronization delay s: any pair $x, y \in Y^s$ (resp. in $P_s \times S_s$) is absorbing, that is $A^*x \cap yA^* \subset M$.

Length distributions

Length distribution of $X \subset A^*$: $u_n = \mathrm{Card}(X \cap A^n)$.

$$\frac{1}{1-u(z)} = \prod_{n \geq 1} \frac{1}{(1-z^n)^{\ell_n(u)}}$$

Theorem (Schützenberger, 1965)

There exists a circular code on k symbols with length distribution $u = (u_n)$ if and only if for all $n \geq 1$ $\ell_n(u)$ is at most equal to the number of primitive necklaces of length n.

Proof: Lazard elimination.

Example: $k = 2$, $u_1 = 1$, $u_2 = 0$, $u_3 = 2$

$$a, \mathbf{b}$$

$$a, \mathbf{ba}, bba, \ldots$$

$$a, baa, bba, \ldots$$

The Franaszek code

$X = \{aaca, aba, aca, , acba, ba, ca, cba\}$ can be obtained as follows (the word to be eliminated is printed in boldface):

$$
\begin{array}{r}
a, b, \mathbf{c} \\
a, b, ca, \mathbf{cb} \\
a, \mathbf{b}, ca, cba \\
\mathbf{a}, ba, ca, cba \\
aaca, aba, aca, acba, ba, ca, cba
\end{array}
$$

Shift of finite type (sft): set of labels of biinfinite paths in a local automaton.
Fondamental example: edge shift of a graph G = set of biinfinite paths in G.

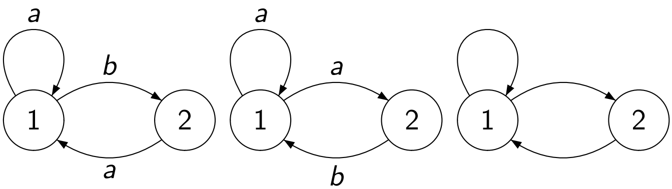


Figure: The golden mean shift

Entropy

For an sft S, denote by v_n the number of distinct blocks of length n of the elements of S.
The entropy of S is

$$h(S) = \lim \frac{1}{n} \log(v_n).$$

The entropy of the golden mean shift is $(1+\sqrt{5})/2$, the dominant eigenvalue of

$$M = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$$

Morphism of sft's: map $\varphi : S \mapsto T$ defined by $y = \varphi(x)$ if

$$y_n = f(x_{n-m} \cdots x_n \cdots x_{n+a})$$

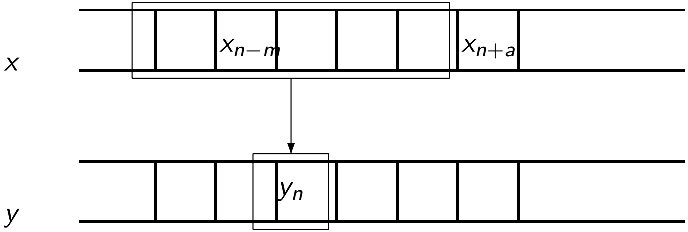


Figure: A sliding block map

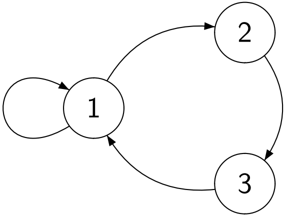
Krieger's embedding theorem

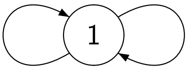
Theorem (Krieger, 1982)

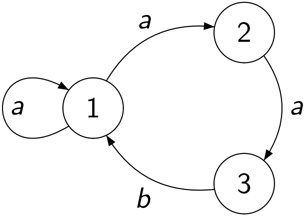
An sft S can be strictly embedded into an sft T if and only if

1. $h(S) < h(T)$
2. $q_n(S) \leq q_n(T)$ *for* $n \geq 1$*, where* $q_n(S)$ *is the number of points of minimal period* n*.*

Link with circular codes: embedding of the edge shift of a bouquet of circles into the full shift.

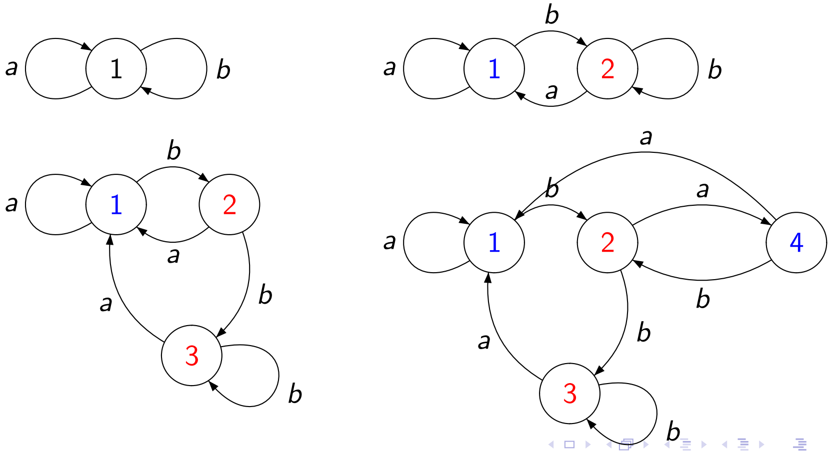






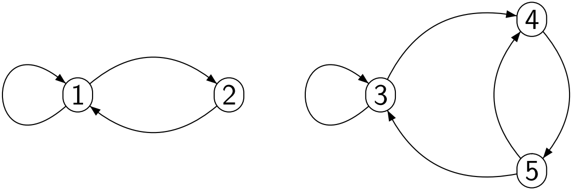
State splitting

Lazard's elimination can be viewed as a sequence of elementary isomorphisms obtained by (input) state-splitting: a state is split into two states with the same output.

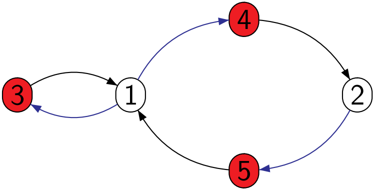


Elementary isomorphisms

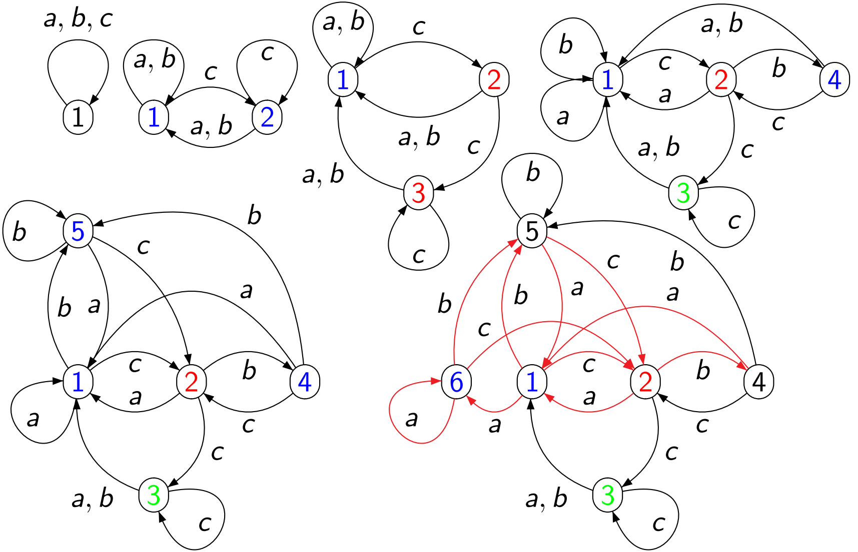
An isomorphism of sft is a composition of elementary isomorphisms obtained by state-splitting (or state-merging).



$$M = RS, \quad R = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, S = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad N = SR$$



Franaszek code



Open question

Is there a proof of Krieger's theorem using an appropriate sequence of splits (elementary isomorphisms)?

The existing proof uses an intricate direct coding.

The masking lemma

For a graph G, denote by X_G the edge shift on G.

Theorem (Nasu, 1988)

Let G and H be graphs. Suppose that X_G embeds into X_H. Then there is a graph K such that $X_K \equiv X_H$ and G is a subgraph of K.

Proof: Let G', H' be the extension of G, H to s-blocks, in such a way that G' is a subgraph of H'. There exists a sequence of graphs

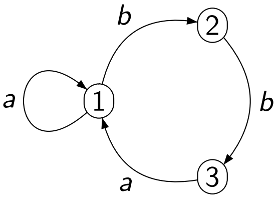
$$G' = G_0, G_1, \ldots, G_n = G$$

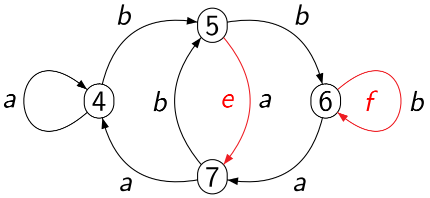
with $G_i \approx G_{i+1}$ and a corresponding sequence

$$H' = H_0, H_1, \ldots, H_n = K$$

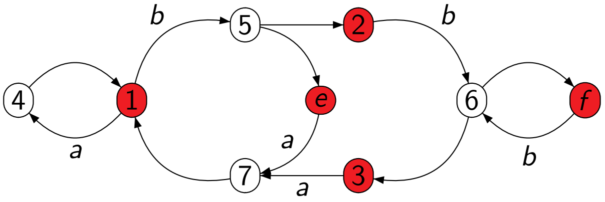
such that G_i is a subgraph of H_i.

Example



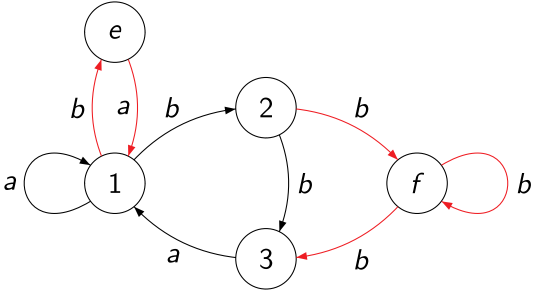


$$M = RS, \quad R = \begin{bmatrix} a & b & 0 & 0 \\ 0 & 0 & b & 0 \\ 0 & 0 & 0 & a \end{bmatrix}, S = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \quad N = SR$$



Theorem

Any local automaton is contained in a local complete automaton.



Question: is the delay preserved?