Open problems on words and groups

Dominique Perrin

15 janvier 2011

1. Bifix codes
   - Internal transformation
   - Ahlswede's conjecture
   - Syntactic groups

2. General codes
   - Finite factorizations
   - Factorization conjecture
   - Noncommutative polynomials

Proposition

Let F be a recurrent set, let $X \subset F$ be a finite F-maximal bifix code and let w be a nonempty word in F. Let $G = Xw^{-1}$, and $D = w^{-1}X$. If $G \neq \emptyset$, $D \neq \emptyset$, and $Gw \cap wD = \emptyset$, then the set

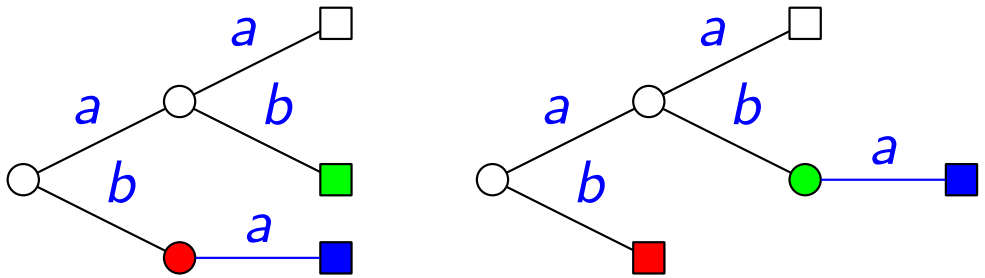
$$Y = (X \cup w \cup (GwD \cap F)) \setminus (Gw \cup wD)$$

is a finite F-maximal bifix code with the same F-degree as X.

The set Y is said to be obtained from X by *internal transformation* (with respect to w).

Example

Let F be the Fibonacci set. The set $X = \{aa, ab, ba\}$ is an F-maximal bifix code of F-degree 2. Then $Y = \{aa, aba, b\}$ is a bifix code of F-degree 2 which is obtained from X by internal transformation with respect to $w = b$. Indeed, here $G = D = \{a\}$, $Gw = \{ab\}$, $wD = \{ba\}$ and $GwD \cap F = \{aba\}$.



Césari's Theorem

Theorem (Césari, 1972)

For any finite maximal bifix code X over A of degree d, there is a sequence of internal transformations which, starting from the code A^d, gives the code X.

This theorem is not true in any recurrent, or even uniformly recurrent set, as shown by the following example.

Example

Let F be the Fibonacci set. The set $X = \{a, bab, baab\}$ is a finite bifix code of F-degree 2. It cannot be obtained by a sequence of internal transformations from the code $A^2 \cap F = \{aa, ab, ba\}$. Indeed, the only internal transformation which can be realized is with respect to $w = b$. The result is $\{aa, aba, b\}$. Next, no internal transformation can be realized from this code.

A modified form

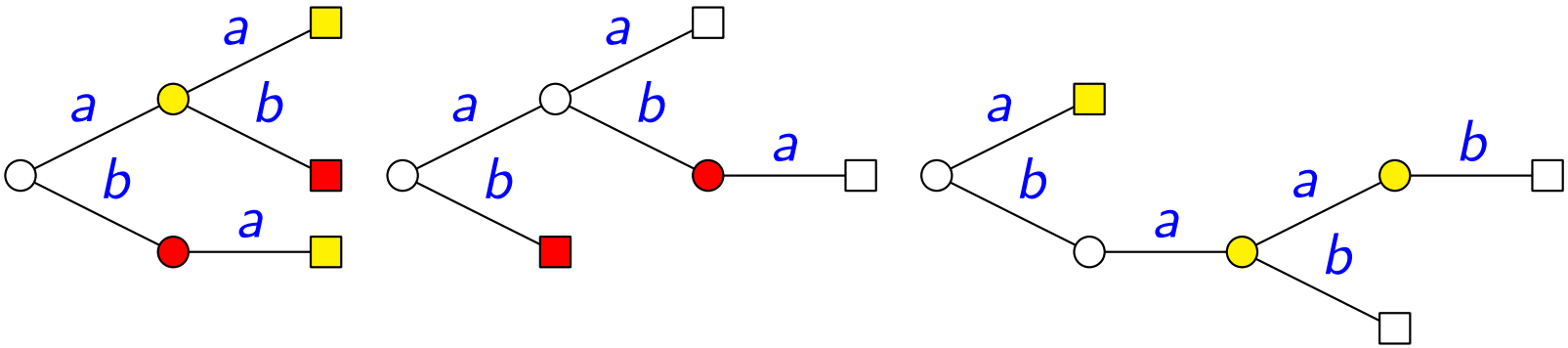
Proposition

Let F be a recurrent set, let $X \subset F$ be a finite F-maximal bifix code and let w be a nonempty word in F. Let $G = Xw^{-1}$, $D = w^{-1}X$, $G_0 = (wD)w^{-1}$, $D_0 = w^{-1}(Gw)$, $G_1 = G \setminus G_0$, $D_1 = D \setminus D_0$. If $G \neq \emptyset$, $D \neq \emptyset$, then the set

$$Y = (X \cup w \cup (G_1 w D_0^* D_1 \cap F)) \setminus (Gw \cup wD)$$

is a finite F-maximal bifix code with the same F-degree as X.

Does the modified form allow one to obtain all finite F-maximal bifix codes of F-degree d starting with the code $A^d \cap F$?


a
a
b
b
a
a
a
b
b
a
a
b
a
b
a
b

For a sequence $(u_n)_{n\geq 1}$ of integers, let $u(z) = \sum_{n\geq 1} u_n z^n$.
Contrary to the case of prefix codes, it is not true that any sequence $(u_n)_{n\geq 1}$ of integers such that $u(1/k) \leq 1$ is the length distribution of a bifix code on k letters. For instance, there is no bifix code on the alphabet $\{a, b\}$ which has the same distribution as the prefix code $\{a, ba, bb\}$.
It is easy to verify that if $u(1/k) \leq 1/2$ there exists a bifix code on a k-letter alphabet with length distribution u_n.

Ahlswede's conjecture

Conjecture (Alswehde, 1996)

For any sequence $(u_n)_{n\geq 1}$ *of integers such that*

$$\sum_{n\geq 1} u_n k^{-n} \leq 3/4$$

there exists a bifix code on an alphabet of k *letters with length distribution* $(u_n)_{n\geq 1}$.

Among the partial results obtained so far, we mention that for $k = 2$, the conjecture holds with $3/4$ replaced by $5/8$, as shown by (Yekhanin, 2004).

N	2			3				4				
	u_1	u_2	$u(1/2)$	u_1	u_2	u_3	$u(1/2)$	u_1	u_2	u_3	u_4	$u(1/2)$
	2	0	1.0000	2	0	0	1.0000	2	0	0	0	1.0000
	1	1	0.7500	1	1	1	0.8750	1	1	1	1	0.9375
				1	0	2	0.7500	1	0	2	1	0.8125
								1	0	1	3	0.8125
								1	0	0	4	0.7500
	0	4	1.0000	0	4	0	1.0000	0	4	0	0	1.0000
				0	3	1	0.8750	0	3	1	0	0.8750
								0	3	0	1	0.8125
				0	2	2	0.7500	0	2	2	2	0.8750
								0	2	1	3	0.8125
								0	2	0	4	0.7500
				0	1	5	0.8750	0	1	5	1	0.9375
								0	1	4	4	1.0000
								0	1	3	5	0.9375

We fix $N \geq 1$ and we order sequences $(u_n)_{1 \leq n \leq N}$ of integers by setting $(u_n) \leq (v_n)$ if and only if $u_n \leq v_n$ for $1 \leq n \leq N$. If $(u_n) \leq (v_n)$ and (v_n) is k-realizable then so is (u_n). We give in Table 1 the values of the maximal 2-realizable sequences for $N \leq 4$. We set $u(z) = \sum u_n z^n$. For each value of N, we list in decreasing lexicographic order the maximal realizable sequence with the corresponding value of the sum $u(1/2) = \sum u_n 2^{-n}$. The distributions with value 1 correspond to maximal bifix codes.

Maximal bifix codes

Since a finite maximal bifix code X is also maximal as a code, its generating series satisfies $f_X(1/k) = 1$, where k is the size of the alphabet. Table 2 lists the length distributions of finite maximal bifix codes of degree $d \leq 4$ over $\{a, b\}$. For each degree, the last column contains the number of bifix codes with this distribution, with a total number of 73 of degree 4. There are 39 of them with $\{a, b\}^3$ as derivative and 34 with one of the two other bifix codes of degree 3.

d	1		2			3					4								
	2	1	0	4	1	0	0	8		1	0	0	0	16					1
											0	0	1	12	4				6
											0	0	2	8	8				6
											0	0	2	9	4	4			8
											0	0	3	5	8	4			6
											0	0	3	6	4	8			4
											0	0	3	6	5	4	4		4
											0	0	4	3	5	8	4		4
						0	1	4	4	2	0	1	0	5	12	4			2
											0	1	0	6	8	8			2
											0	1	0	6	9	4	4		4
											0	1	0	7	5	8	4		4
											0	1	0	7	6	5	4	4	2
											0	1	0	8	2	9	4	4	2
											0	1	1	3	9	8	4		4

Groups of codes

Proposition

Let X be a thin maximal prefix code. If the group $G = G(X)$ admits an imprimitivity equivalence θ, then there exists a decomposition of X into

$$X = Y \circ Z$$

such that $G(Y) = G^\theta$ and $G(Z) = G_\theta$.

Problem : is this proposition holds for arbitrary thin maximal codes.

Next, let $X \subset A^+$ be a finite code with n elements and let $\mathcal{A} = (Q, 1, 1)$ be a trim unambiguous automaton recognizing X^*. Let $\varphi = \varphi_{\mathcal{A}}$ and let $M = \varphi(A^*)$. Let e be an idempotent in the transition monoid of the automaton $\mathcal{A}$ and let H be the $\mathcal{H}$-class of e. Schützenberger (1979) has proved that either $\varphi^{-1}(H)$ is cyclic or the group G_e has degree at most $2n$. This bound can be reduced to n by using the critical factorization theorem. It is conjectured that actually, the degree of G_e is at most $n-1$ if $\varphi^{-1}(H)$ is not cyclic. This is known to be true if X is prefix (Perrin, Rindone 2003).

The rank conjecture

The minimal rank of a group is the minimal cardinality of a generating set.

Conjecture

Let X be a finite bifix code and let G be a permutation group of degree d and minimal rank k. If G is a syntactic group of X, then $\mathrm{Card}(X) \geq (k-1)d + 1$.

Theorem

Let X be a finite maximal bifix code of degree $d \geq 4$. Then $G(X)$ is not a Frobenius group.

It is not known whether this theorem holds more generally for finite maximal prefix codes. For example, it is not known if there exists a finite maximal prefix code X such that $G(X)$ is the dihedral group D_5.

Finite factorizations

Let $p, q \geq 0$ be two integers. A code $X \subset A^*$ is said to be (p, q)-limited if for any sequence $u_0, u_1, \ldots, u_{p+q}$ of words in A^*, the assumptions

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

imply

$$u_p \in X^* .$$

Given a factorization $A^* = \underline{X_n^*}\ \underline{X_{n-1}^*} \cdots \underline{X_1^*}$ with n factors, are the codes X_i always limited ? This is true if the factorization is obtained by iterating bisections. It is true for factorizations with up to four factors by a result of Krob (1987).

Conjecture

A conjecture in relation with factorizations is the following. If $\underline{A^*} = \underline{M_1} \cdots \underline{M_n}$ *where* $M_1, \ldots, M_n$ *are submonoids, then the* M_i *are free submonoids.*

This is known to hold up to $n = 4$ (Krob, 1987).

Factorization conjecture

A code X is positively factorizing if there exists a pair (P, S) of sets such that each word $w \in A^*$ factorizes uniquely into

$$w = sxp \tag{2}$$

with $p \in P$, $s \in S$, $x \in X^*$.
The factorization conjecture states that any finite maximal code is positively factorizing.
The commutative equivalence conjecture states that any finite maximal code is commutatively prefix. The factorization conjecture implies the commutative equivalence conjecture. There are relations between the factorization conjecture and factorizations of cyclic groups.

Noncommutative polynomials

Let K be a field and let A be an alphabet. A subring R of $K\langle A\rangle$ is free if it is isomorphic to $K\langle B\rangle$ for some alphabet B. A subring R of $K\langle A\rangle$ is called an anti-ideal if for any $u \in K\langle A\rangle$ and nonzero $v, w \in R$, $uv, wu \in R$ implies $u \in R$. By a theorem of Kolotov (1978), a free subring of $K\langle A\rangle$ is an anti-ideal. Thus the subring generated by a submonoid M of A^* is an anti-ideal if and only if it is free. Indeed, if $K\langle M\rangle$ is an anti-ideal, then M is stable and therefore is free. This is not true for arbitrary subrings of $K\langle A\rangle$ (Cohn, 1985) It is not known whether the property that $K\langle Y\rangle$ is free, for a finite set Y of $K\langle A\rangle$, is decidable.