Vietnam Journal of Mathematics **32**:3 (2004) 345–359

Vietnam Journal
of
MATHEMATICS

Codes Concerning Roots of Words

Kieu Van Hung[1], Phan Trung Huy[2], and Do Long Van[3]

[1] *Hanoi Pedagogical University No. 2, Phuc Yen, Vinh Phuc, Vietnam*
[2] *Hanoi Polytechnic University, Dai Co Viet Street, Hanoi, Vietnam*
[3] *Hanoi Institute of Math., 18 Hoang Quoc Viet Road, 10307 Hanoi, Vietnam*

Received December 17, 2003

Abstract. We consider some classes of codes defined by binary relations concerning roots of words. An extension of the embedding schema, introduced in 1998 by the last of the authors, is proposed. This leads to positive solution of the embedding problem, in the finite case, for the classes of codes under consideration.

1. Preliminaries

For many classes of codes, a slight application of the Zorn's lemma shows that every code in a class is included in a code maximal in the same class (not necessarily maximal as a code). So, in certain sense, these maximal codes completely characterize the whole class of codes.

For such a given class C of codes, a natural question is whether every code X satisfying some property $\mathfrak{p}$ (usually, the finiteness or the regularity) is included in a code Y maximal in C which still has the property $\mathfrak{p}$. This problem, which we call the *embedding problem* for the class C, attracts a lot of interests. Unfortunately, this problem was solved only for several cases by means of different combinatorial techniques.

In 1998, a general embedding schema for the classes of codes, which can be defined by *length-increasing transitive* binary relations, was proposed [6]. This allows to solve positively, in a unified way, for both the finite and the regular case, the embedding problem for many classes of codes well-known as well as new (see [6, 7]).

Our aim in this paper is to extend the mentioned above embedding schema to include more classes of codes. For this we replace in the schema the requirement "length-increasing" by a weaker one, namely the "f-increasing". Based on the extended embedding schema we show that the embedding problem is solved positively for some new classes of codes, whose definitions deal with roots of words. These classes are all subclasses of prefix or suffix codes. Also, some properties of these codes are considered.

We now recall some notions, notations and facts, which will be used in the sequel. Let A throughout be a finite alphabet and A^* the set of all the words over A. The empty word is denoted by 1 and A^+ stands for $A^* - \{1\}$. The number of all the occurrences of letters in a word u is the *length* of u, denoted by $|u|$. Any subset of A^* is a *language* over A. A language X over A is a *code* if the equality

$$x_1 x_2 \dots x_n = y_1 y_2 \dots y_m,$$

where $x_1, \dots, x_n, y_1, \dots, y_m \in X$, implies $n = m$ and $x_1 = y_1, \dots, x_n = y_n$. A code over A is *maximal* if it is not properly contained in any code over A. Let C be a class of codes over A and $X \in C$. We say that the code X is *maximal in* C if it is not properly contained in any code in C. For the background of the theory of codes we refer to [1, 3, 4].

Binary relations appeared to be a good tool to introduce new classes of codes [2, 5 - 7]. Let $\prec$ be a binary relation on A^* and X be a non-empty subset of A^*. The language X is said to be an *independent set w.r.t.* $\prec$ if there do not exist any different elements $x, y \in X$ such that $x \prec y$. We say that a class C of codes is *defined by* $\prec$ if these codes are exactly the independent sets w.r.t. $\prec$. Then, we denote the class C by $C_\prec$. Very often the relation $\prec$ characterizes some property α of words, then we write $\prec_\alpha$ instead of $\prec$ and, for the sake of simplicity, C_α stands for $C_{\prec_\alpha}$. The relation $\prec$ is said to be *length-increasing* if for any $u, v \in A^*$ with $u \neq v$, $u \prec v$ implies $|u| < |v|$. We denote by $\preceq$ the reflexive closure of $\prec$, i.e. for any $u, v \in A^*, u \preceq v$ if and only if $u \prec v$ or $u = v$.

The following binary relations on A^*, as easily verified, are transitive (except for $\prec_b$) and length-increasing [6, 7].

$u \prec_p v \Leftrightarrow v = ux$, with $x \neq 1$;
$u \prec_s v \Leftrightarrow v = xu$, with $x \neq 1$;
$u \prec_b v \Leftrightarrow (u \prec_p v) \vee (u \prec_s v)$;
$u \prec_{p.i} v \Leftrightarrow v = xuy$, with $y \neq 1$;
$u \prec_{s.i} v \Leftrightarrow v = xuy$, with $x \neq 1$;
$u \prec_i v \Leftrightarrow v = xuy$, with $xy \neq 1$;
$u \prec_{p.h} v \Leftrightarrow \exists n \geq 1 : u = u_1 \dots u_n \wedge v = x_0 u_1 \dots u_n x_n$, with $x_1 \dots x_n \neq 1$;
$u \prec_{s.h} v \Leftrightarrow \exists n \geq 1 : u = u_1 \dots u_n \wedge v = x_0 u_1 \dots u_n x_n$, with $x_0 \dots x_{n-1} \neq 1$;
$u \prec_h v \Leftrightarrow \exists n \geq 1 : u = u_1 \dots u_n \wedge v = x_0 u_1 x_1 \dots u_n x_n$, with $x_0 \dots x_n \neq 1$;
$u \prec_{p.si} v \Leftrightarrow \exists w \in A^* : w \prec_p v \wedge u \preceq_h w$;
$u \prec_{s.si} v \Leftrightarrow \exists w \in A^* : w \prec_s v \wedge u \preceq_h w$;
$u \prec_{si} v \Leftrightarrow \exists w \in A^* : w \prec_i v \wedge u \preceq_h w$;
$u \prec_{p.scpi} v \Leftrightarrow (\exists v' : v' \prec_p v)(\exists v'' \in \sigma(v')) : u \preceq_h v''$;
$u \prec_{s.scpi} v \Leftrightarrow (\exists v' : v' \prec_s v)(\exists v'' \in \sigma(v')) : u \preceq_h v''$;
$u \prec_{scpi} v \Leftrightarrow (\exists v' : v' \prec_i v)(\exists v'' \in \sigma(v')) : u \preceq_h v''$;
$u \prec_{p.spi} v \Leftrightarrow (\exists v' : v' \prec_p v)(\exists v'' \in \pi(v')) : u \preceq_h v''$;
$u \prec_{s.spi} v \Leftrightarrow (\exists v' : v' \prec_s v)(\exists v'' \in \pi(v')) : u \preceq_h v''$;

$u \prec_{spi} v \Leftrightarrow (\exists v' : v' \prec_i v)(\exists v'' \in \pi(v')) : u \preceq_h v''$;
$u \prec_{scp} v \Leftrightarrow \exists v' \in \sigma(v) : u \prec_h v'$;
$u \prec_{sp} v \Leftrightarrow \exists v' \in \pi(v) : u \prec_h v'$,
where $\pi(v)$ and $\sigma(v)$ denote the sets of all permutations and all cyclic permutations of the word v, respectively.

The listed above relations define corresponding classes of codes which are denoted by and named, respectively, the classes C_p of *prefix codes*, C_s of *suffix codes*, C_b of *bifix codes*, $C_{p.i}$ of *p-infix codes*, $C_{s.i}$ of *s-infix codes*, C_i of *infix codes*, $C_{p.h}$ of *p-hypercodes*, $C_{s.h}$ of *s-hypercodes*, C_h of *hypercodes*, $C_{p.si}$ of *p-subinfix codes*, $C_{s.si}$ of *s-subinfix codes*, C_{si} of *subinfix codes*, $C_{p.scpi}$ of *p-sucyperinfix codes*, $C_{s.spci}$ of *s-sucyperinfix codes*, C_{spci} of *sucyperinfix codes*, $C_{p.spi}$ of *p-superinfix codes*, $C_{s.spi}$ of *s-superinfix codes*, C_{spi} of *superinfix codes*, C_{scp} of *sucypercodes* and C_{sp} of *supercodes*. By virtue of the embedding schema in [6] all these classes, except for the class C_b of bifix codes, have been proved to have positive solution for the embedding problem in both the finite and regular cases [6, 7].

The following fact has been shown in [7].

Fact 1.1. *Let $\prec_1$ and $\prec_2$ be binary relations on A^*. Then $C_{\prec_1 \cup \prec_2} = C_{\prec_1} \cap C_{\prec_2}$.*

Let $u \in A^+$, the word u is called *primitive* if $u = r^e$ for some $r \in A^+$ implies $e = 1$. The unique primitive word r such that $u = r^e$ for some integer e is called the *root* of u. We shall denote by Q the set of all primitive words on A. For any $i \geq 2$, set $Q^{(i)} = \{r^i \mid r \in Q\}$, $Q^{(1)} = Q \cup \{1\}$. Then, obviously $Q^{(i)} \cap Q^{(j)} = \emptyset$ for $i \neq j$ and $A^* = \bigcup_{i=1}^{+\infty} Q^{(i)}$ (see [1, 4]). The following result, which will be used in Sec. 4, is due to D. Borwein (see [4, p. 8]).

Proposition 1.2. *Let $u \in A^+, u \neq a^n, a \in A$. Then one of the words ua and u must be primitive.*

2. New Classes of Codes

We introduce in this section some new classes of codes. They are all defined by binary relations concerning roots of words.

Definition 2.1. *Given an integer $n \geq 1$ and $u \in A^+$. We call root of u with threshold n (n-root of u, for short) the word $r_n(u)$ defined as follows*

$$r_n(u) = \begin{cases} r, & \text{if } u = r^e, r \in Q, 1 \leq e \leq n; \\ u, & \text{if } u = r^e, r \in Q, e > n. \end{cases}$$

By convention, $r_n(1) = 1$. This operation can be extended to languages in a standard way, namely $r_n(X) = \{r_n(u) \mid u \in X\}$.

Definition 2.2. *Let $\Omega = \{p, s, b, p.i, s.i, i, p.si, s.si, si, p.scpi, s.scpi, scpi, p.spi, s.spi, spi, p.h, s.h, h, scp, sp\}$. To every $\alpha \in \Omega$ and every positive integer n we* *associate a binary relation on A^*, denoted $\prec_{n.\alpha}$, which is given by*

$$u \prec_{n.\alpha} v \Leftrightarrow r_n(u) \preceq_\alpha r_n(v).$$

We denote by $C_{n.\alpha}$ the class of all the independent sets on A^* w.r.t. the relation $\prec_{n.\alpha}$. All such classes, as we shall see later (Theorems 2.5 and 2.6), are subclasses of prefix codes or suffix codes. In other words, $C_{n.\alpha}$ is nothing but the class of codes defined by the relation $\prec_{n.\alpha}$. Evidently, $C_{1.\alpha} \equiv C_\alpha$ for all $\alpha \in \Omega$.

If α is p, for example, then the relation $\prec_{n.p}$ defines the class $C_{n.p}$, whose members are called *r_n-prefix codes*, etc.

Example 2.3. Consider $X = \{b^k a, b^{k+1}\}$ $(k \geq 1)$ over $A = \{a, b\}$. Since $r_k(b^{k+1}) = b^{k+1} \npreceq_p b^k a = r_k(b^k a)$, X is a r_k-prefix code. But it is not a r_{k+1}-prefix code because $r_{k+1}(b^{k+1}) = b \prec_p b^k a = r_{k+1}(b^k a)$.

Example 2.4. Let $A = \{a, b\}$ and $k \geq 1$. Consider the following sets

$$X_1 = \{aaba, (aba)^k\},\ X_2 = \{bbaba, (bab)^k\},\ X_3 = \{abaab, (bab)^k\},$$
$$X_4 = \{abaabb, (bab)^k\},\ X_5 = ba^+b,\ X_6 = \{baab, (aba)^k\},$$
$$X_7 = \{a^k, ba\},\ X_8 = \{abba, (aba)^k\},\ X_9 = \{baab, (abb)^k\}.$$

It is not difficult to check that $X_5 \in C_{n.si} \cap C_{n.scpi} \cap C_{n.spi}$, and for any $n \geq k$ hold the following

$X_1 \in C_{n.p.i} - C_{n.i}, X_1^R \in C_{n.s.i} - C_{n.i}; X_2 \in C_{n.b} - C_{n.p.i} \cup C_{n.s.i};$
$X_3 \in C_{n.p.spi} - C_{n.si}; X_3^R \in C_{n.s.spi} - C_{n.si}; X_4 \in C_{n.p.i} \cup C_{n.s.i} - C_{n.si};$
$X_7 \in C_{n.p.h} - C_{n.h}, X_7^R \in C_{n.s.h} - C_{n.h}; X_8 \in C_{n.si} - C_{n.p.h} \cup C_{n.s.h};$
$X_6 \in C_{n.si} - C_{n.p.spi} \cup C_{n.s.spi}; X_9 \in C_{n.scpi} \cap C_{n.spi} \cap C_{n.h} - C_{n.scp} - C_{n.sp}.$

It appears that the classes $C_{n.p}$, $C_{n.s}$ and $C_{n.b}$ constitute infinite hierarchies on n w.r.t. inclusion. More precisely we have

Theorem 2.5. *For any integers $m, k \geq 1$, $m > k$ implies $C_{m.p} \subset C_{k.p}$, $C_{m.s} \subset C_{k.s}$ and $C_{m.b} \subset C_{k.b}$. In particular, for any $n \geq 1$, $C_{n.p} \subset C_{1.p} \equiv C_p$, $C_{n.s} \subset C_{1.s} \equiv C_s$ and $C_{n.b} \subset C_{1.b} \equiv C_b$.*

Proof. We treat only the case of r_n-prefix codes. For the other cases the argument is similar. Let k, m be positive integers such $m > k$. We first show that if $u \prec_{k.p} v$ then either $u \prec_{m.p} v$ or $v \prec_{m.p} u$. Indeed, by definition, $r_k(u) \preceq_p r_k(v)$. Since $m > k$, it follows that $r_m(u) \preceq_p r_k(u)$. By the transitivity of $\preceq_p$, we obtain $r_m(u) \preceq_p r_k(v)$. But $r_m(v) \preceq_p r_k(v)$ because $m > k$. Therefore, either $r_m(u) \preceq_p r_m(v)$ or $r_m(v) \preceq_p r_m(u)$, that means $u \prec_{m.p} v$ or $v \prec_{m.p} u$. Now, given a set X in $C_{m.p}$. By definition, $u \not\prec_{m.p} v$ for all $u, v \in X$ with $u \neq v$. If $u \prec_{k.p} v$ then, by the above, either $u \prec_{m.p} v$ or $v \prec_{m.p} u$, a contradiction with $X \in C_{m.p}$. Thus $u \not\prec_{k.p} v$ and therefore $X \in C_{k.p}$. Hence $C_{m.p} \subseteq C_{k.p}$. The Example 2.3 shows the strictness of the inclusion. ∎

Note that for the classes $C_{n.\alpha}, \alpha \in \Omega - \{p, s, b\}$, there do not exist similar assertions as in Theorem 2.5. Indeed, consider the set $X = \{baa, (aab)^i\}$ $(i \geq 2)$ over $A = \{a, b\}$. A direct verification shows that $X \in C_{m.\alpha}$ but $X \notin C_{k.\alpha}$, for any $\alpha \in \Omega - \{p, s, b\}$ and for any positive integers m, k such that $m \geq i > k$.

Relationship between the classes of codes under consideration is given below.

Theorem 2.6. *For every $n \geq 1$, we have the following*

(i) $C_{n.b} \subset C_{n.p}, C_{n.b} \subset C_{n.s}, C_{n.b} = C_{n.p} \cap C_{n.s}$;

(ii) $C_{n.i} \subset C_{n.p.i}$, $C_{n.i} \subset C_{n.s.i}$, $C_{n.i} = C_{n.p.i} \cap C_{n.s.i}$, $C_{n.i} \subset C_{n.b}$, $C_{n.p.i} \subset C_{n.p}$, $C_{n.s.i} \subset C_{n.s}$;

(iii) $C_{n.si} \subset C_{n.p.si}$, $C_{n.si} \subset C_{n.s.si}$, $C_{n.si} = C_{n.p.si} \cap C_{n.s.si}$, $C_{n.si} \subset C_{n.i}$, $C_{n.p.si} \subset C_{n.p.i}$, $C_{n.s.si} \subset C_{n.s.i}$;

(iv) $C_{n.scpi} \subset C_{n.p.scpi}$, $C_{n.scpi} \subset C_{n.s.scpi}$, $C_{n.scpi} = C_{n.p.scpi} \cap C_{n.s.scpi}$, $C_{n.scpi} \subset C_{n.si}$, $C_{n.p.scpi} \subset C_{n.p.si}$, $C_{n.s.scpi} \subset C_{n.s.si}$;

(v) $C_{n.spi} \subset C_{n.p.spi}$, $C_{n.spi} \subset C_{n.s.spi}$, $C_{n.spi} = C_{n.p.spi} \cap C_{n.s.spi}$, $C_{n.spi} \subset C_{n.scpi}$, $C_{n.p.spi} \subset C_{n.p.scpi}$, $C_{n.s.spi} \subset C_{n.s.scpi}$;

(vi) $C_{n.h} \subset C_{n.p.h}$, $C_{n.h} \subset C_{n.s.h}$, $C_{n.h} \subset C_{n.si}$, $C_{n.p.h} \subset C_{n.p.si}$, $C_{n.s.h} \subset C_{n.s.si}$, $C_{n.h} = C_{n.p.h} \cap C_{n.s.h} = C_{n.p.h} \cap C_{n.s.si} = C_{n.p.si} \cap C_{n.s.h}$;

(vii) $C_{n.sp} \subset C_{n.scp} \subset C_{n.h}$, $C_{n.scp} \subset C_{n.scpi}$, $C_{n.sp} \subset C_{n.spi}$.

Proof. We prove only the item (ii). For the remaining items the argument is similar. Suppose that $u \prec_{n.p.i} v$, which means $r_n(u) \preceq_{p.i} r_n(v)$. It follows that $r_n(u) \preceq_i r_n(v)$, and therefore $u \prec_{n.i} v$. Thus $C_{n.i} \subseteq C_{n.p.i}$. Similarly, $C_{n.i} \subseteq C_{n.s.i}$. The sets X_1 and X_1^R in Example 2.4 show that the inclusions are strict. In the same way, we obtain $C_{n.i} \subseteq C_{n.b}$, $C_{n.p.i} \subseteq C_{n.p}$ and $C_{n.s.i} \subseteq C_{n.s}$. The set X_2 in Example 2.4 proves the strictness of the inclusions. Next, by definition of the relations $\prec_{n.p.i}, \prec_{n.s.i}$ and $\prec_{n.i}$, we have $\prec_{n.i} = \prec_{n.p.i} \cup \prec_{n.s.i}$. Hence, by Fact 1.1, $C_{n.i} = C_{n.p.i} \cap C_{n.s.i}$. ∎

By virtue of Theorem 2.6 the relative positions of the classes of codes under consideration can be illustrated in the Fig. 1, where the arrow $\to$ stands for a strict inclusion. It is worthy to note that if we restrict ourselves to considering only one-letter alphabets then all the classes of codes represented in the Fig. 1 coincide.

As known in [1, 2, 4], the classes C_p, C_s, C_b, $C_{p.i}$, $C_{s.i}$, C_i and C_h are closed under concatenation. Now we shall show that the classes $C_{p.si}$, $C_{s.si}$ and C_{si} are also closed under concatenation whereas all the other classes under consideration are not.

Theorem 2.7. *The following assertions hold true*

(i) *The classes $C_{p.si}$ of p-subinfix codes, $C_{s.si}$ of s-subinfix codes and C_{si} of subinfix codes are closed under concatenation.*

(ii) *If $n > 1$ and $\alpha \in \Omega$ or $n = 1$ and $\alpha \in \Omega - \{p, s, b, p.i, s.i, i, p.si, s.si, si, h\}$, the class $C_{n.\alpha}$ is not closed under concatenation.*

Proof. (i) We treat only the class $C_{p.si}$ of p-subinfix codes. For the other cases the argument is similar. Let X and Y be two p-subinfix codes over A and $Z = XY$.

Suppose $z, w \in Z$ such that z is a subword of w. Then, on one hand, $z = xy$, $w = x'y'$ for some $x, x' \in X$ and $y, y' \in Y$. On the other hand, $z = z_1 \dots z_k$ and $w = w_0 z_1 w_1 \dots z_k w_k$ with $z_1, \dots, z_k \in A, w_0, w_1, \dots, w_k \in A^*$. Since $z = z_1 \dots z_k$, there exists $t, 1 \le t < k$, such that $x = z_1 \dots z_t, y = z_{t+1} \dots z_k$.

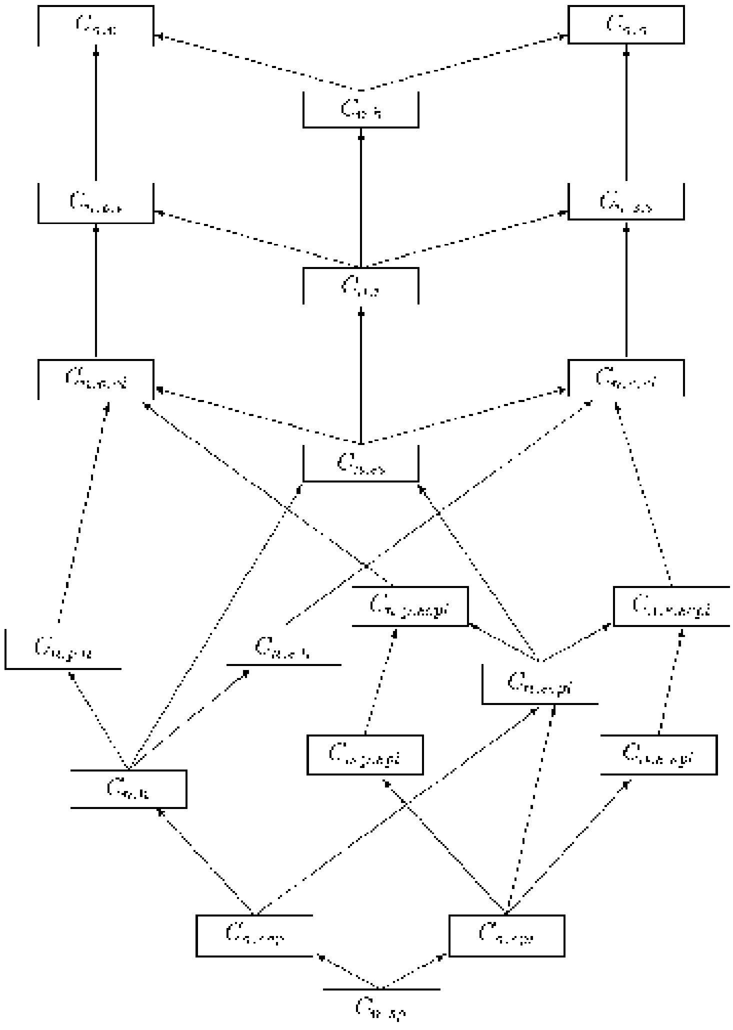

Figure 1. Relative positions of the classes $C_{n.\alpha}$ of codes with n fixed

From

$$w_0 z_1 w_1 \dots z_t w_t z_{t+1} \dots z_k w_k = x'y'$$

it follows that either $x' = w_0 z_1 w_1 \dots z_t$ or $x' = w_0 z_1 w_1 \dots z_m w'_m$ $(m < t,$

$w'_m \prec_p w_m$) because otherwise we have $x \prec_{p.si} x'$, a contradiction. Thus $y' = uz_{t+1}w_{t+1} \ldots z_k w_k$, $u \in A^*$. Since $y' \in Y$, we must have $w_k = 1$. This means that z may not be a subword of a proper prefix of w. Thus, $Z = XY$ must be a p-subinfix code.

(ii) Let's consider the subsets $X = \{a, b\}$, $Y = \{a^2b, ab^3\}$, $Z = \{a, ba\}$, $X' = \{bab, ba^2b\}$, $Y' = \{baba, a^3b^3\}$, $Z' = \{baba, a^2b^3\}$ over the alphabet $A = \{a, b\}$. It is easy to check that the following things hold true, which prove the assertion required.

$X \in C_{n.\alpha}, X^2 \notin C_{n.\alpha}$, for all $\alpha \in \Omega$ and $n > 1$;
$X, Y \in C_{sp} \subset C_{scp}$, $YX \notin C_{sp} \cup C_{scp}$;
$X, Z \in C_{p.h}, X, Z^R \in C_{s.h}$, $XZ \notin C_{p.h}, Z^R X \notin C_{s.h}$;
$X, X' \in C_{spi} = C_{p.spi} \cap C_{s.spi}$, $XX' \notin C_{p.spi} \cup C_{s.spi}$;
$Y', Z', {Y'}^R, {Z'}^R \in C_{scpi}$, $Y'Z' \notin C_{p.scpi}$, ${Z'}^R {Y'}^R \notin C_{s.scpi}$. ∎

3. Extended Embedding Schema

Our aim in this section is to give an extension of the embedding schema in [6]. For this we need some more definitions and notations.

Definition 3.1. *Let $f : A^* \to \mathbb{N}$ be a function from A^* into the set $\mathbb{N}$ of natural numbers. Let $\prec$ be a binary relation on A^*. The relation $\prec$ is said to be f-increasing if, for any $u, v \in A^*$, $(u \prec v) \wedge (v \not\prec u)$ implies $f(u) < f(v)$. We say that f is a function with finite condition on inverse images (fc-function, for short) if for any $k \in \mathbb{N}$, $f^{-1}(k)$ is a finite set, or equivalently, for any $X \subseteq A^*$,* $\mathrm{Card}\, f(X) < +\infty$ *implies* $\mathrm{Card}\, X < +\infty$, *where* $\mathrm{Card}\, Y$ *denotes the cardinality of the set Y.*

Definition 3.2. *Let $\prec$ be a transitive binary relation on A^*. We denote by $\simeq_\prec$ the equivalence relation on A^* given by*

$$u \simeq_\prec v \Leftrightarrow (u \preceq v) \wedge (v \preceq u)$$

Let X be a subset of A^* and $w \in X$. We denote by $[w]_{\prec,X}$, or simply $[w]_X$ if there is no risque of confusion, the equivalence class of w restricted on X, i.e.

$$[w]_{\prec,X} = [w]_{\simeq_\prec} \cap X = \{u \in X \mid (u \preceq w) \wedge (w \preceq u)\}$$

The family of these sets constitutes a partition of the set X, denoted by $[X]_\prec$ or simply $[X]$. Any function $\varphi : [X]_\prec \to X$ mapping every set $[w]_{\prec,X}$ into an element of it is called a choice function of $[X]_\prec$.

Given a binary relation $\prec$ on A^* and $u, v \in A^*$. We say that u *depends* on v if either $u \preceq v$ or $v \preceq u$ holds. Otherwise, u is *independent* of v. These notions are extended to subsets of words in a standard way. Namely, a word u is dependent on a subset X of words if it depends on some word in X. Otherwise, u is independent of X. For simplicity, the following notations will be used in the sequel.

$$u \preceq X \rightleftharpoons \exists v \in X : u \preceq v; \qquad X \preceq u \rightleftharpoons \exists v \in X : v \preceq u.$$

Defintion 3.3. *Let $\prec$ be a transitive binary relation on A^*. To every subset X in A^* one associates the sets D_X and I_X consisting of all the elements dependent on and independent of X respectively, i.e.*

$$D_X = \{u \in A^* \mid u \preceq X \vee X \preceq u\};$$
$$I_X = A^* - D_X.$$

Notice that, by definition, $X \subseteq D_X$. Next, for any choice function φ of $[I_X]_\prec$ we put

$$K_X = \varphi([I_X]_\prec);$$
$$L_X = \{u \in K_X \mid \exists v \in K_X, v \neq u : v \prec u\};$$
$$R_X = K_X - L_X.$$

When there can not be confusion, for the sake of simplicity, we write D, I, K, L, R instead.

Lemma 3.4. *Let $\prec$ be a transitive binary relation on A^*. Let $w \in I_X - R_X$ such that $X \cup R_X \cup \{w\}$ be an independent set w.r.t. $\prec$. Then, for any $u \in [w]_{I_X}$ we have $u \notin X \cup R_X$ and $X \cup R_X \cup \{u\}$ is also an independent set w.r.t. $\prec$ on A^*.*

Proof. The fact that $u \notin X \cup R_X$ is trivial. Suppose u depends on $X \cup R_X$, i.e. there exists $x \in X \cup R_X$ such that either $x \preceq u$ or $u \preceq x$. Since $u \in [w]_{I_X}$, we have $u \preceq w$ and $w \preceq u$. Therefore, either $x \preceq u \preceq w$ or $w \preceq u \preceq x$. By the transitivity of $\prec$, either $x \preceq w$ or $w \preceq x$, a contradiction. So $X \cup R_X \cup \{u\}$ is an independent set w.r.t. $\prec$ on A^*. ■

We are now in a position to formulate the main result of this paper. When we take as f the length-function, i.e. $f(u) = |u|$, we obtain again the embedding schema established in [6].

Theorem 3.5. *Let $\prec$ be a transitive f-increasing binary relation on A^* which defines the class $C_\prec$ of codes. Then, for any code X in $C_\prec$ and for any choice function φ of $[I_X]_\prec$, we have*

(i) *$X \cup R_X$ is a maximal code in $C_\prec$;*

(ii) *If moreover X is finite, f is a fc-function and the relation $\prec$ satisfies the condition*

$$\exists k \geq 1 \forall u, v \in A^+ : (f(v) \geq f(u) + k) \wedge (u \not\prec v)$$
$$\Rightarrow \exists w : (f(w) \geq f(u)) \wedge (w \prec v) \wedge (v \not\prec w) \qquad (*)$$

then R_X is finite and $\max_f R_X \leq \max_f X + k - 1$, where $\max_f Y$ stands for $\max\{f(y) \mid y \in Y\}$.

Proof. (i) First we show that $X \cup R_X$ is an independent set w.r.t. $\prec$. Assume the contrary, there would exist $u, v \in X \cup R_X$ with $u \neq v$ such that $u \prec v$. Because X is an independent set w.r.t. $\prec$, it is impossible that both u and v are in X. Since R_X contains only elements independent of X, it is impossible also that one of these two elements u and v is in X and the other in R_X. So $u, v \in R_X$, which contradicts $u \prec v$. Thus, $X \cup R_X$ is an independent set w.r.t. $\prec$, i.e. it is a code in $C_\prec$.

We now prove that $X \cup R_X$ is a maximal independent set w.r.t. $\prec$. Suppose the contrary that there exists $w \notin X \cup R_X$ such that $X \cup R_X \cup \{w\}$ is an independent set w.r.t. $\prec$. If $w \in D_X$ then w depends on X, a contradiction. Thus $w \in I_X$. Choose $v_0 = \varphi([w]_{I_X}) \in K_X$. By Lemma 3.4, $v_0 \notin X \cup R_X$ and $X \cup R_X \cup \{v_0\}$ is an independent set. Therefore $v_0 \in L_X$. By the definition of L_X, there exists $v_1 \in K_X$ with $v_1 \neq v_0$ such that $v_1 \prec v_0$. If $v_1 \in R_X$ then $X \cup R_X \cup \{v_0\}$ is not an independent set, a contradiction. Hence $v_1 \in L_X$. Then, there must exist $v_2 \in K_X$ with $v_2 \neq v_1$ such that $v_2 \prec v_1$. The transitivity of $\prec$ implies $v_2 \prec v_0$. If $v_2 \in R_X$ then $v_2 \neq v_0$ and therefore $X \cup R_X \cup \{v_0\}$ is not an independent set, a contradiction. Hence $v_2 \in L_X$. Continuing this argument we obtain finally an infinite sequence of elements $v_i \in L_X \subseteq K_X$, $i \geq 0$, such that

$$\cdots \prec v_{i+1} \prec v_i \prec \cdots \prec v_1 \prec v_0$$

with $v_i \neq v_{i+1}$ for all $i \geq 0$. We have also $v_i \not\prec v_{i+1}$. Indeed, if this is not the case then $v_i, v_{i+1} \in [z]_{I_X}$ for some $z \in I_X$, which implies $v_i, v_{i+1} \in [z]_{I_X} \cap K_X$ with $v_i \neq v_{i+1}$, a contradiction with the fact that $\mathrm{Card}\,([z]_{I_X} \cap K_X) = 1$. Since $\prec$ is f-increasing, $f(v_{i+1}) < f(v_i)$ for all $i \geq 0$. This means the chain

$$\cdots < f(v_{i+1}) < f(v_i) < \cdots < f(v_1) < f(v_0)$$

is infinite, which is impossible because $f(v_0)$ is finite. Thus $X \cup R_X$ must be a maximal independent set w.r.t. $\prec$, i.e. it is a maximal code in $C_\prec$.
(ii) Suppose X is a finite code in $C_\prec$ with $m = \max_f X$, f is a fc-function and $\prec$ satisfies the condition $(*)$. We shall prove that for any $v \in K_X$ with $f(v) \geq m + k$ there exists $w \in K_X$ with $w \neq v$ such that $w \prec v$, and therefore $v \in L_X$. This assertion means that for all $v \in R_X$, $f(v) \leq m + k - 1$, i.e. $\max_f R_X \leq \max_f X + k - 1$. Since f is a fc-function, it follows the finiteness of R_X, which completes the proof. We prove now the formulated above assertion. Indeed, let u be an element in X with maximal $f(u)$, i.e. $f(u) = m$. Then $f(v) \geq f(u) + k$. Since every element in I_X, in K_X in particular, is independent of X, we have $u \not\prec v$. By the condition $(*)$, there exists w' such that $f(w') \geq f(u)$, $w' \prec v$ and $v \not\prec w'$. As $v \in K_X$, it follows that $w' \notin X$. Suppose $w' \in (D_X - X)$, then there exists $x \in X$ such that either $x \prec w'$ or $w' \prec x$. If $x \prec w'$ then, by the transitivity of $\prec$, $x \prec v$, which contradicts the hypothesis $v \in K_X$. Thus $w' \prec x$ and $x \not\prec w'$. Because $\prec$ is f-increasing, it follows that $f(w') < f(x)$, which contradicts the fact that $f(w') \geq f(u) \geq f(x)$. Thus $w' \in I_X$. Then, for all $z \in [w']_{I_X}$, $z \prec v$ because $z \preceq w' \prec v$, and $z \neq v$ because $v \not\prec w'$. Take $w = \varphi([w']_{I_X})$ we have $w \in K_X$, $w \prec v$ and $w \neq v$. Thereby the assertion has been proved. ■

4. An Embedding Theorem

In this section we apply Theorem 3.5 to show that the embedding problem for the classes of codes introduced in Sec. 2 is solved positively in the finite case. For proving this result we need some lemmas.

To every integer $n \geq 1$, we associate a function $\ell_n : A^* \to \mathbb{N}$ mapping every word u into the length of its n-root, $\ell_n(u) = |r_n(u)|$. In particular, ℓ_1 is nothing but the length-function, $\ell_1(u) = |u|$.

Lemma 4.1. *For any integer $n \geq 1$, ℓ_n is a fc-function.*

Proof. It suffices to prove that for any $m \in \mathbb{N}$, the set ${\ell_n}^{-1}(m)$ is finite. For $m = 0$ it is true trivially because ${\ell_n}^{-1}(0) = \{1\}$. Suppose $m \neq 0$ and let $u \in {\ell_n}^{-1}(m)$, $u = r^e$ with $r \in Q$ and $e \geq 1$. By the definition of $r_n(u)$, it follows that either $|u| = m$ or $|u| = |r^e| = e.|r_n(u)| \leq n.m$ according as $e > n$ or $e \leq n$. Hence ${\ell_n}^{-1}(m)$, being a subset of $\{u \in A^* \mid |u| \leq n.m\}$, is finite. ∎

Lemma 4.2. *For any integer $n \geq 1$ and α in Ω, the relation $\prec_{n.\alpha}$ is ℓ_n-increasing. Also this relation is transitive, except for $\alpha = b$.*

Proof. The transitivity of $\prec_{n.\alpha}$, except for $\alpha = b$, is evident by definition. Suppose $u, v \in A^*$ such that $u \prec_{n.\alpha} v$ and $v \not\prec_{n.\alpha} u$. By the definition of $\prec_{n.\alpha}$, $r_n(u) \preceq_\alpha r_n(v)$ and $r_n(v) \not\preceq_\alpha r_n(u)$. Thus $r_n(u) \preceq_\alpha r_n(v)$ and $r_n(v) \neq r_n(u)$, and therefore $r_n(u) \prec_\alpha r_n(v)$. So $\ell_n(u) < \ell_n(v)$, which means that $\prec_{n.\alpha}$ is ℓ_n-increasing. ∎

In general, $\prec_{n.\alpha}$ with $\alpha \in \Omega$ is not length-increasing. For example, for any $n \geq k > 1$, we have $r_n(a^k) = a \preceq_\alpha a = r_n(a)$. Thus $a^k \prec_{n.\alpha} a$ whereas $|a^k| > |a|$.

Lemma 4.3. *For any integer $n \geq 1$ and α in Ω, the relation $\prec_{n.\alpha}$ satisfies the condition $(*)$ of Theorem 3.5 with $k = n$. More concretely, for any $u, v \in A^+$ such that $|r_n(v)| \geq |r_n(u)| + n$ and $u \not\prec_{n.\alpha} v$, there exists w such that $|r_n(w)| \geq |r_n(u)|$, $w \prec_{n.\alpha} v$ and $v \not\prec_{n.\alpha} w$.*

Proof. Put $\Gamma = \{s, s.i, s.si, s.scpi, s.spi, s.h\}$ and let $\gamma \in \Omega - \Gamma$. Assume that $u, v \in A^+$ such that $|r_n(v)| \geq |r_n(u)| + n$ and $u \not\prec_{n.\gamma} v$. We will show that there exists w such that $|r_n(w)| \geq |r_n(u)|$, $w \prec_{n.\gamma} v$ and $v \not\prec_{n.\gamma} w$.

For $n = 1$, the assertion has been proved in [6]. Suppose $n > 1$ and $r_n(v) = x^t$, with $x \in Q$, $t \geq 1$. Since $r_n(v) \in Q + \bigcup_{j \geq n+1} Q^{(j)}$, only the following two cases are possible:

Case 1. $t = 1$. Since $|x| = |r_n(v)| \geq |r_n(u)| + n > 2$, the word x may be written as $x = a_1 \ldots a_k$ with $a_1, \ldots, a_k \in A$ and $k > 2$. Put $w = a_1 \ldots a_{k-1}$ and let $w = y^s$, with $y \in Q$, $s \geq 1$.

If either $s = 1$ or $s \geq n + 1$ then $r_n(w) = w$, therefore $|r_n(w)| = |w| = |r_n(v)| - 1 \geq |r_n(u)| + n - 1 > |r_n(u)|$. Since $r_n(w) = w \prec_p x = r_n(v)$, it follows that $r_n(w) \prec_\gamma r_n(v)$. Thus $w \prec_{n.\gamma} v$ and $v \not\prec_{n.\gamma} w$.

If $1 < s \leq n$ then $r_n(w) = y$. So, if $|y| = 1$ then $s = |w| = |r_n(v)| - 1 \geq$

$|r_n(u)|+n-1 \geq n$. Thus $s = n$ and therefore we have $|r_n(u)|+n \leq |r_n(v)| = |x| = k = s+1 = n+1$. Hence $|r_n(u)| = 1 = |r_n(w)|$. As $r_n(w) = y \prec_p w \prec_p r_n(v)$, it follows that $r_n(w) \prec_\gamma r_n(v)$. Thus $w \prec_{n.\gamma} v$ and $v \not\prec_{n.\gamma} w$. Suppose now $|y| > 1$ and take $w' = a_1 \ldots a_{k-2}$. We have $|r_n(w')| = |r_n(v)| - 2 \geq |r_n(u)| + n - 2 \geq |r_n(u)|$. It is impossible that $w' = a^q$ for some $q \geq 1$ and $a \in A$. Since $w = w'a_{k-1} \notin Q$, by Proposition 1.2, $w' \in Q$. Hence $r_n(w') = w' \prec_\gamma r_n(v)$. Thus $w' \prec_{n.\gamma} v$ and $v \not\prec_{n.\gamma} w'$.

Case 2. $t \geq n+1$. Suppose $|x| = 1$ and put $w = x^{t-1}$. Since $r_n(w) \preceq_p w \prec_p r_n(v)$, we obtain $r_n(w) \prec_\gamma r_n(v)$. It follows that $w \prec_{n.\gamma} v$ and $v \not\prec_{n.\gamma} w$. If $t > n+1$, i.e. $t-1 \geq n+1$ then $r_n(w) = w = x^{t-1}$ and therefore $|r_n(w)| = |x^{t-1}| = t-1 = |x^t| - 1 = |r_n(v)| - 1 \geq |r_n(u)| + n - 1 > |r_n(u)|$. If $t = n+1$, i.e. $t-1 = n$ then $|r_n(w)| = 1 = (n+1) - n = |r_n(v)| - n \geq |r_n(u)|$. Suppose now $|x| > 1$ and take $w = x^{t-1}a_1 \ldots a_{k-1}$. It is impossible that $w = a^q$ for some $q \geq 1$ and $a \in A$. Since $x^t = wa_k \notin Q$, by Proposition 1.2, $w \in Q$. Hence $r_n(w) = w \prec_p r_n(v)$. Thus $r_n(w) \prec_\gamma r_n(v)$. So $w \prec_{n.\gamma} v$, $v \not\prec_{n.\gamma} w$, and $|r_n(w)| = |r_n(v)| - 1 \geq |r_n(u)| + n - 1 > |r_n(u)|$.

For the relations $\prec_{n.\gamma}$ with $\gamma \in \Gamma$, the proof is quite similar where $\prec_s$ is used instead of $\prec_p$. ■

Lemma 4.4. *For any $n \geq 1, \alpha \in \Omega - \{b\}, X \in C_{n.\alpha}$, $w \in I_X$, and $w = r^e$ with $r \in Q, e \geq 1$, we have*

$$[w]_{\prec_{n.\alpha}, I_X} = \begin{cases} \{r, r^2, \ldots, r^n\} & \text{if } e \leq n; \\ \{w\} & \text{otherwise.} \end{cases}$$

Proof. Let $n \geq 1$ and $\alpha \in \Omega - \{b\}$. Let $w \in A^+$ and $w = r^e$ with $r \in Q, e \geq 1$. We have obviously

$$\begin{aligned} [w]_{\prec_{n.\alpha}, A^+} &= \{u \in A^+ \mid (u \preceq_{n.\alpha} w) \wedge (w \preceq_{n.\alpha} u)\} \\ &= \{u \in A^+ \mid (r_n(u) \preceq_\alpha r_n(w)) \wedge (r_n(w) \preceq_\alpha r_n(u))\} \\ &= \{u \in A^+ \mid r_n(u) = r_n(w)\} \\ &= \begin{cases} \{r, r^2, \ldots, r^n\} & \text{if } e \leq n; \\ \{w\} & \text{otherwise.} \end{cases} \end{aligned}$$

Next, we claim that, if $r \in I_X \cap Q$ then $r^i \in I_X$ for all $i = 1, \ldots, n$. Indeed, for $n = 1$ this is true trivially. Suppose $n > 1$ and there exists a word $r^e \in D_X$, with $e \in \{2, \ldots, n\}$. Then, there is a word $x \in X$ such that either $r^e \preceq_{n.\alpha} x$ or $x \preceq_{n.\alpha} r^e$. Since $r_n(r) = r_n(r^e) = r$, we have $r \preceq_{n.\alpha} r^e$, $r^e \preceq_{n.\alpha} r$. By the transitivity of $\prec_{n.\alpha}$, it follows that either $r \preceq_{n.\alpha} x$ or $x \preceq_{n.\alpha} r$. This means $r \in D_X$, which contradicts the hypothesis that $r \in I_X$. Thus the claim has been proved, and thereby, the expression of $[w]_{\prec_{n.\alpha}, I_X}$ is obtained immediately from the established above expression of $[w]_{\prec_{n.\alpha}, A^+}$. ■

Lemma 4.5. *For any $n \geq 1, \alpha \in \Omega - \{b\}$ and $X \in C_{n.\alpha}$, there exists a choice function φ_0 of $[I_X]_{\prec_{n.\alpha}}$ such that* $\max R_X = \max_{\ell_n} R_X$.

Proof. Let $w \in I_X$ and $w = r^e$ with $r \in Q, e \geq 1$. By Lemma 4.4, as a choice function of $[I_X]_{\prec_{n.\alpha}}$ we may take the function φ_0 defined by

$$\varphi_0([w]_{\prec_{n.\alpha}, I_X}) = \begin{cases} r, & \text{if } [w]_{\prec_{n.\alpha}, I_X} = \{r, r^2, \ldots, r^n\}; \\ w, & \text{otherwise.} \end{cases}$$

Then, obviously $K_X \subseteq Q + \bigcup_{i>n} Q^{(i)}$. Therefore, for any u in K_X, in R_X in particular, we have $r_n(u) = u$, hence $\ell_n(u) = |u|$. Therefore $\max R_X = \max_{\ell_n} R_X$. ■

Theorem 4.6. *For any $n \geq 1$ and $\alpha \in \Omega - \{b\}$, every finite code X in the class $C_{n.\alpha}$ is included in a finite maximal code Y in the same class, which contains X, namely $Y = X \cup R_X$ with $\max_{\ell_n} R_X \leq \max_{\ell_n} X + n - 1$. Moreover R_X can be chosen such that $\max R_X \leq \max_{\ell_n} X + n - 1$ and hence $\max Y \leq \max\{\max X, \max_{\ell_n} X + n - 1\}$, where $\max Z$ denotes the maximal length of the words in Z.*

Proof. By Lemma 4.2, the relation $\prec_{n.\alpha}$ is transitive and ℓ_n-increasing. By Lemma 4.1, ℓ_n is a fc-function. By Lemma 4.3, the relation $\prec_{n.\alpha}$ satisfies the condition $(*)$ in Theorem 3.5 with $k = n$. By Theorem 3.5, the set $Y = X \cup R_X$ is a finite maximal code in $C_{n.\alpha}$ and $\max_{\ell_n} R_X \leq \max_{\ell_n} X + n - 1$. Taking φ_0 as the choice function of $[I_X]_{\prec_{n.\alpha}}$, by Lemma 4.5, it follows that $\max R_X = \max_{\ell_n} R_X \leq \max_{\ell_n} X + n - 1$, hence $\max Y \leq \max\{\max X, \max_{\ell_n} X + n - 1\}$. ■

5. Corollaries

Theorem 4.6 permits us to obtain algorithms to construct, for every finite code X in a class $C_{n.\alpha}$, $\alpha \in \Omega - \{b\}$, finite maximal codes in the same class which contain X. As examples of application, we exhibit below such algorithms for the class $C_{n.p}$ of r_n-prefix codes and the class $C_{n.p.scpi}$ of r_n-p-sucyperinfix codes. Algorithms for the other classes can be obtained in a similar way.

For simplicity of notation, in the sequel we shall write $\overline{r}_n(w)$ instead of $[w]_{\prec_{n.\alpha}, A^*}$. We denote by $A^{[k]}$ the set of all the words in A^* whose lengths are less than or equal to k. Also we put $Q^{>k} = \bigcup_{i>k} Q^{(i)}$.

Corollary 5.1. *Let X be a finite r_n-prefix code over the alphabet A and $m = \max_{\ell_n} X + n - 1$. Then a finite maximal r_n-prefix code Y containing X can be computed by the following formulas*

$$\begin{aligned} &Y = X \cup R_X,\ R_X = K_X - L_X,\ L_X = K_X \cap \big(K_X A^+ \cap A^{[m]}\big), \\ &K_X = I_X \cap (Q \cup Q^{>n}),\ I_X = A^{[m]} - D_X, \\ &D_X = \big(\overline{r}_n(r_n(X)(A^*)^{-1}) \cap A^{[m]}\big) \\ &\qquad \cup \big(\overline{r}_n(r_n(X)A^+ \cap A^{[m]} \cap (Q \cup Q^{>n})) \cap A^{[m]}\big). \end{aligned}$$

Proof. (sketch) By Theorem 4.6, the set $Y = X \cup R_X$ is a finite maximal r_n-prefix code containing X and $\max R_X \leq m$. So, R_X can be obtained by computing step by step the sets D_X, I_X, K_X, L_X, and in every step we restrict to considering only the words in $A^{[m]}$. Notice that $K_X = \varphi_0([I_X]_{\prec_{n.p}}) = I_X \cap (Q \cup Q^{>n})$. ■

Example 5.2. Consider the language $X = \{a^2, (ba^2b)^2\}$ over the alphabet $A = \{a, b\}$. Since $r_2(X) = \{a, ba^2b\}$ is a prefix code, X is a r_2-prefix code and $\max_{\ell_2} X = 4$. We now compute R_X, using the formulas in Corollary 5.1 with $m = \max_{\ell_2} X + 2 - 1 = 5$.

$$r_2(X)(A^*)^{-1} = \{1, a, b, ba, ba^2, ba^2b\};$$

$$\overline{r}_2(r_2(X)(A^*)^{-1}) = \{1, a, a^2, b, b^2, ba, (ba)^2, ba^2, (ba^2)^2, ba^2b, (ba^2b)^2\};$$

$$\overline{r}_2(r_2(X)(A^*)^{-1}) \cap A^{[5]} = \{1, a, a^2, b, b^2, ba, (ba)^2, ba^2, ba^2b\};$$

$$\begin{aligned} r_2(X)A^+ \cap A^{[5]} \cap (Q \cup Q^{>2}) &= (aA + aA^2 + aA^3 + aA^4 + ba^2bA) \cap (Q \cup Q^{>2}) \\ = \{&\underline{a}^2, ab, a^3, a^2b, aba, ab^2, a^4, a^3b, a^2ba, a^2b^2, aba^2, (\underline{ab})^2, ab^2a, ab^3, \\ &a^5, a^4b, a^3ba, a^3b^2, a^2ba^2, a^2bab, a^2b^2a, a^2b^3, aba^3, aba^2b, \\ &a(ba)^2, (ab)^2b, ab^2a^2, ab^2ab, ab^3a, ab^4, ba^2ba, ba^2b^2\} \cap (Q \cup Q^{>2}) \\ = \{&ab, a^3, a^2b, aba, ab^2, a^4, a^3b, a^2ba, a^2b^2, aba^2, ab^2a, ab^3, \\ &a^5, a^4b, a^3ba, a^3b^2, a^2ba^2, a^2bab, a^2b^2a, a^2b^3, aba^3, aba^2b, \\ &a(ba)^2, (ab)^2b, ab^2a^2, ab^2ab, ab^3a, ab^4, ba^2ba, ba^2b^2\}; \end{aligned}$$

$$\begin{aligned} &\overline{r}_2(r_2(X)A^+ \cap A^{[5]} \cap (Q \cup Q^{>2})) \cap A^{[5]} = \{ab, a^3, a^2b, aba, ab^2, a^4, a^3b, \\ &\quad a^2ba, a^2b^2, aba^2, (\underline{ab})^2, ab^2a, ab^3, a^5, a^4b, a^3ba, a^3b^2, a^2ba^2, a^2bab, a^2b^2a, \\ &\quad a^2b^3, aba^3, aba^2b, a(ba)^2, (ab)^2b, ab^2a^2, ab^2ab, ab^3a, ab^4, ba^2ba, ba^2b^2\}; \end{aligned}$$

$$\begin{aligned} D_X = \{&1, a, a^2, b, b^2, ba, (ba)^2, ba^2, ba^2b, ab, a^3, a^2b, aba, ab^2, a^4, a^3b, a^2ba, \\ &a^2b^2, aba^2, (\underline{ab})^2, ab^2a, ab^3, a^5, a^4b, a^3ba, a^3b^2, a^2ba^2, a^2bab, a^2b^2a, a^2b^3, \\ &aba^3, aba^2b, a(ba)^2, (ab)^2b, ab^2a^2, ab^2ab, ab^3a, ab^4, ba^2ba, ba^2b^2\}; \end{aligned}$$

$$\begin{aligned} I_X = A^{[5]} - D_X = \{&bab, b^2a, b^3, ba^3, bab^2, b^2a^2, b^2ab, b^3a, b^4, ba^4, ba^3b, \\ &(ba)^2a, (ba)^2b, bab^2a, bab^3, b^2a^3, b^2a^2b, b(ba)^2, b^2ab^2, b^3a^2, b^3ab, b^4a, b^5\}; \end{aligned}$$

$$K_X = I_X \cap (Q \cup Q^{>2}) = I_X$$

$$\begin{aligned} L_X = K_X \cap (K_X A^+ \cap A^{[5]}) = \{&bab^2, b^2a^2, b^2ab, b^3a, b^4, ba^4, ba^3b, (ba)^2a, \\ &(ba)^2b, bab^2a, bab^3, b^2a^3, b^2a^2b, b(ba)^2, b^2ab^2, b^3a^2, b^3ab, b^4a, b^5\}; \end{aligned}$$

$$R_X = K_X - L_X = \{bab, b^2a, b^3, ba^3\}.$$

Thus the set $Y = X \cup R_X = \{a^2, bab, b^2a, b^3, ba^3, (ba^2b)^2\}$ is a finite maximal r_2-prefix code containing X.

Remark. When replacing an element u in R_X by an element v in $\overline{r}_2(u)$ we obtain another finite maximal r_2-prefix code containing X. For example, the languages $\{a^2, (bab)^2, b^2a, b^3, ba^3, (ba^2b)^2\}$ and $\{a^2, bab, (b^2a)^2, b^3, (ba^3)^2, (ba^2b)^2\}$ are such codes.

For any set X we denote by $\mathcal{P}(X)$ the family of all subsets of X. Recall that a *substitution* is a mapping f from B into $\mathcal{P}(C^*)$, where B and C are alphabets.

When $f(b)$ is a singleton for all $b \in B$ it induces a *homomorphism* from B^* into C^*. Let # be a new letter not being in A. Put $A_\# = A \cup \{\#\}$. Let us consider the regular substitutions S_1, S_2 and the homomorphism h defined as follows

$S_1 : A \to \mathcal{P}(A^*_\#)$, where $S_1(a) = \{a, \#\}$ for all $a \in A$

$S_2 : A_\# \to \mathcal{P}(A^*)$, with $S_2(\#) = A^+$ and $S_2(a) = \{a\}$ for all $a \in A$

$h : A^*_\# \to A^*$, with $h(\#) = 1$ and $h(a) = a$ for all $a \in A$

Factually, the substitution S_1 will be used to mark the occurrences of letters to be deleted from a word. The homomorphism h realizes the deletion by replacing # by the empty word. The inverse homomorphism h^{-1} "chooses" in a word the positions where the words of A^+ may be inserted, while S_2 realizes the insertions by replacing # with A^+.

We present now an embedding algorithm for the case of finite r_n-p-sucyperinfix codes.

Corollary 5.3. *Let X be a finite r_n-p-sucyperinfix code over the alphabet A and $m = \max_{\ell_n} X + n - 1$. Then a finite maximal r_n-p-sucyperinfix code Y containing X may be computed by the following formulas*

$$Y = X \cup R_X,\ R_X = K_X - L_X,$$

$$L_X = K_X \cap\ S_2(h^{-1}(\sigma(K_X)) \cap A_\#^{[m-1]}\{\#\}) \cap A^{[m]},$$

$$K_X = I_X \cap (Q \cup Q^{>n}),\ I_X = A^{[m]} - D_X,$$

$$D_X = (\overline{r}_n(\sigma(h(S_1(r_n(X)) \cap A^*_\#\{\#\})) \cup X) \cap A^{[m]})$$

$$\cup (\overline{r}_n(S_2(h^{-1}(\sigma(r_n(X))) \cap A_\#^{[m-1]}\{\#\}) \cap A^{[m]} \cap (Q \cup Q^{>n})) \cap A^{[m]}).$$

Example 5.4. Let us consider the language $X = \{(ab)^2, a^2b\}$ over the alphabet $A = \{a, b\}$. It is not difficult to check that, for any $n \geq 2$, this language is an r_n-p-sucyperinfix code, not being a p-sucyperinfix code. Since $\max_{\ell_n} X = 3$, we may compute R_X by the formulas in Corollary 5.3 with $m = 3 + n - 1$. Let us take $n = 2$, then $m = 4$, and R_X can be computed step by step as follows.

$r_2(X) = \{ab, a^2b\}$;

$S_1(r_2(X)) \cap A^*_\#\{\#\}) = \{a\#, \#\#, a^2\#, \#a\#, a\#\#, \#\#\#\}$;

$\sigma(h(S_1(r_2(X)) \cap A^*_\#\{\#\})) \cup X = \{1, a, a^2\} \cup \{(ab)^2, a^2b\}$;

$\overline{r}_2(\sigma(h(S_1(r_2(X)) \cap A^*_\#\{\#\})) \cup X) \cap A^{[4]} = \{1, a, a^2, ab, (ab)^2, a^2b\}$;

$\sigma(r_2(X)) = \{ab, ba, a^2b, aba, ba^2\}$;

$h^{-1}(\sigma(r_2(X))) \cap A_\#^{[3]}\{\#\} = \{ab\#, \#ab\#, ab\#\#, a\#b\#,$

$ba\#, \#ba\#, ba\#\#, b\#a\#, a^2b\#, aba\#, ba^2\#\}$;

$S_2(h^{-1}(\sigma(r_2(X))) \cap A_\#^{[3]}\{\#\}) \cap A^{[4]} \cap (Q \cup Q^{>2}) = (abA + AabA + abA^2$

$+ aAbA + baA + AbaA + baA^2 + bAaA + a^2bA + abaA + ba^2A) \cap (Q \cup Q^{>2})$

$= \{aba, ab^2, a^2ba, a^2b^2, (\underline{ba})^2, bab^2, aba^2, (\underline{ab})^2,$

$= \{ab^2a, ab^3, ba^2, bab, b^2a^2, b^2ab, ba^3, ba^2b\} \cap (Q \cup Q^{>2})$

$= \{aba, ab^2, a^2ba, a^2b^2, bab^2, aba^2, ab^2a, ab^3, ba^2, bab, b^2a^2, b^2ab, ba^3, ba^2b\}$;

$$\begin{aligned}
&\overline{r}_2(S_2(h^{-1}(\sigma(r_2(X))) \cap A_{\#}^{[3]}\{\#\}) \cap A^{[4]} \cap (Q \cup Q^{>2})) \cap A^{[4]} \\
&= \{aba, ab^2, a^2ba, a^2b^2, bab^2, aba^2, ab^2a, ab^3, ba^2, bab, b^2a^2, b^2ab, ba^3, ba^2b\}; \\
&D_X = \{1, a, a^2, ab, (ab)^2, a^2b, aba, ab^2, a^2ba, a^2b^2, \\
&D_X = \{bab^2, aba^2, ab^2a, ab^3, ba^2, bab, b^2a^2, b^2ab, ba^3, ba^2b\}; \\
&I_X = A^{[4]} - D_X = \{b, ba, \underline{b^2}, a^3, b^2a, b^3, a^4, a^3b, (\underline{ba})^2, b^3a, b^4\}; \\
&K_X = I_X \cap (Q \cup Q^{>2}) = \{b, ba, a^3, b^2a, b^3, a^4, a^3b, b^3a, b^4\}; \\
&h^{-1}(\sigma(K_X)) \cap A_{\#}^{[3]}\{\#\} = \{b\#, \#b\#, b\#\#, \#\#b\#, \#b\#\#, b\#\#\#, \#ab\#, \\
&ab\#, ab\#\#, a\#b\#, ba\#, \#ba\#, ba\#\#, b\#a\#, a^3\#, ab^2\#, bab\#, b^2a\#, b^3\#\}; \\
&S_2(h^{-1}(\sigma(K_X)) \cap A_{\#}^{[3]}\{\#\}) \cap A^{[4]} = \{ba, aba, ab^2, b^2a, b^3, ba^2, bab, \\
&\quad a^2ba, a^2b^2, ab^2a, ab^3, bab^2, b^3a, b^4, aba^2, b^2a^2, b^2ab, ba^3, ba^2b, a^4, a^3b\}; \\
&L_X = K_X \cap S_2(h^{-1}(\sigma(K_X)) \cap A_{\#}^{[3]}\{\#\}) \cap A^{[4]} = \{ba, b^2a, b^3, a^4, a^3b, b^3a, b^4\}; \\
&R_X = K_X - L_X = \{b, a^3\}.
\end{aligned}$$

Thus the set $Y = X \cup R_X = \{b, a^3, (ab)^2, a^2b\}$ is a finite maximal r_2-p-sucyperinfix code containing X. If consider X as a r_3-p-sucyperinfix code instead, by a similar computation, we obtain the set $Y' = \{b, a^3b, a^4, (ab)^2, a^2b\}$ which is a finite maximal r_3-p-sucyperinfix code containing X.

References

1. J. Berstel and D. Perrin, *Theory of Codes*, Academic Press, New York, 1985.
2. M. Ito, H. Jürgensen, H. Shyr, and G. Thierrin, Outfix and infix codes and related classes of languages, *J. Computer and System Sciences* **43** (1991) 484–508.
3. H. Jürgensen and S. Konstatinidis, Codes, G. Rozenberg and A. Salomaa (Eds.), *Handbook of formal languages*, Springer, Berlin, 1997, 511–607.
4. H. Shyr, *Free Monoids and Languages*, Hon Min Book Company, Taichung, 1991.
5. H. Shyr and G. Thierrin, Codes and binary relations, Lecture notes 586 Séminarie d'Algèbre, Paul Dubreil, Paris, Springer-Verlag, 1975-1976, 180–188.
6. D. L. Van, *Embedding Problem for Codes Defined by Binary Relations*, Preprint 98/A22, Institute of Mathematics, Hanoi, 1998.
7. D. L. Van and K. V. Hung, An approach to the embedding problem for codes defined by binary relations, *J. Automata, Languages and Combinatorics*, 2004, (Submitted).