RESEARCH ARTICLE

On the structure of some group codes

Dongyang Long*

Communicated by B. M. Schein

In this paper, we study the following problem: Which characteristics does a code $C$ possess when the syntactic monoid $\text{syn}(C^*)$ of the star closure $C^*$ of $C$ is a group? For a code $C$, if the syntactic monoid $\text{syn}(C^*)$ is a group, then we call $C$ a group code. This definition of a group code is different from the one in [1] (see [1], 46-47). Schützenberger had characterized the structure of finite group codes and had proved that $C$ is a finite group code if and only if $C$ is a full uniform code (see [5], [8]). For $k$-prefix and $k$-suffix codes with $k \geq 2$, $k$-infix, $k$-outfix, $p$-infix, $s$-infix, right semaphore codes and left semaphore codes, etc., we obtain similar results. It is proved that the above mentioned codes are group codes if and only if they are uniform codes.

1. Basic notions and notation

We first introduce the necessary concepts and some notation. For additional details and definitions, see the references, in particular [1], [5], and [7]. Let $X$ be an alphabet. Then $X^*$ denotes the free monoid generated by $X$, that is, the set of all words over $X$, including the empty word $\lambda$, and $X^+ = X^* \setminus \{\lambda\}$. For $w \in X^*$, by $|w|$ we denote the length of $w$. A language $L$ over $X$ is a set $L \subseteq X^*$. With any language $L \subseteq X^*$, one associates its principal congruence $P_L$ and its syntactic monoid $\text{syn}(L) = X^*/P_L$, where

$$u \equiv v(P_L) \iff ((\forall x, y \in X^*) xuy \in L \iff xvy \in L).$$

For $w \in X^*$, by $\bar{w}$ we denote the $P_L$-class of $w$.

A language $L \subseteq X^*$ is said to be a code over $X$ if the submonoid $L^*$ of $X^*$ generated by $L$ is freely generated by $L$. If $P$ is any property of languages, we call a code $C$ a $P$-code if $C$ possesses the property $P$. If $C$ is a $P$-code and for every $u \in X^*$, $C \cup \{u\}$ is not a $P$-code, then $C$ is a maximal $P$-code. For the sake of simplicity, in this paper we assume that $X$ is the least alphabet for a code, that is, let $C$ be a code over $X$, then $X^*aX^* \cap C \neq \emptyset$ for every $a \in X$.

Definition 1. (See [2], [3], [7].) Let $X$ be an alphabet and $k$ be a given positive integer. A language $C \subseteq X^*$ is

(a) a $k$-prefix code if for all $x_1, \ldots, x_k, y_1, \ldots, y_k \in X^*$, $x_1 \ldots x_k \in C$ and $x_1y_1x_2 \ldots x_ky_k \in C$ together imply $y_1 \ldots y_k = 1$;

(b) a $k$-suffix code if for all $x_1, \ldots, x_k, y_1, \ldots, y_k \in X^*$, $x_1 \ldots x_k \in C$ and $y_1x_2y_1 \ldots y_kx_k \in C$ together imply $y_1 \ldots y_k = 1$;

* The author would like to thank the referee who helped the author with revising the paper. Thanks are also due to Prof. Boris M. Schein for his valuable suggestions and kind help.
(c) a $k$-infix code if for all $x_1, \ldots, x_k$, $y_0, \ldots, y_k \in X^*$, $x_1 \ldots x_k \in C$ and $y_0 x_1 y_1 \ldots x_k y_k \in C$ together imply $y_0 \ldots y_k = 1$;

(d) a $k$-outfix code if for all $x_0, \ldots, x_k$, $y_1, \ldots, y_k \in X^*$, $x_0 \ldots x_k \in C$ and $x_0 y_1 x_1 \ldots y_k x_k \in C$ together imply $y_1 \ldots y_k = 1$;

(e) a hypercode if for any natural number $n$ and all $x_1, \ldots, x_n$, $y_0, \ldots, y_n \in X^*$, $x_1 \ldots x_k \in C$ and $y_0 x_1 y_1 \ldots x_n y_n \in C$ together imply $y_0 y_1 \ldots y_k = 1$;

(f) a full uniform code if there exists some integer $m \geq 0$ such that $C = X^m$.

By $P_k(X)$, $S_k(X)$, $I_k(X)$, $O_k(X)$, $H(X)$ and $FUF(X)$ we denote the classes of $k$-prefix codes, $k$-suffix codes, $k$-infix codes, $k$-outfix codes, hypercodes and full uniform codes over $X$, respectively. In particular, $P(X) = P_1(X)$, $S(X) = S_1(X)$, $I(X) = I_1(X)$, $O(X) = O_1(X)$ are the classes of prefix, suffix, infix, and outfix codes, respectively.

Definition 2. (See [1], [2].) Let $X$ be an alphabet. A language $C \subseteq X^*$ is

(a) a bifix code if $C$ is both a prefix and a suffix code;

(b) reflective if for all $u, v \in X^*$, $uv \in C$ imply $vu \in C$;

(c) a $p$-infix code if for all $x, u, y \in X^*$, $xuy \in C$ and $u \in C$ together imply $y = 1$;

(d) an $s$-infix code if for all $x, u, y \in X^*$, $xuy \in C$ and $u \in C$ together imply $x = 1$;

(e) a right semaphore code if $C$ is a prefix code satisfying $X^*C \subseteq CX$;

(f) a left semaphore code if $C$ is a suffix code satisfying $CX \subseteq X^*C$.

By $B(X)$, $RE(X)$, $PI(X)$, $SI(X)$, $RSP(X)$, and $LSP(X)$ we denote the classes of bifix, reflective, $p$-infix, $s$-infix, right semaphore and left semaphore codes over $X$, respectively.

According to the results of [3] and Figure 1 in [4], the relations between the classes of codes defined by Definitions 1 and 2 are shown in Figures 1 and 2, respectively.

Remark 1. In [3], we proved the following:

1. $S_{k+1}(X) \subseteq P_k(X)$, $P_{k+1}(X) \subseteq S_k(X)$, $O_{k+1}(X) \subseteq I_k(X)$, $I_{k+1}(X) \subseteq O_k(X)$.

2. $P_{k+1}(X) \subseteq I_k(X)$, $P_{k+1}(X) \subseteq O_k(X)$, $S_{k+1}(X) \subseteq I_k(X)$, $S_{k+1}(X) \subseteq O_k(X)$.

3. $I_k(X) \subseteq P_k(X)$, $O_k(X) \subseteq P_k(X)$, $I_k(X) \subseteq S_k(X)$, $O_k(X) \subseteq S_k(X)$.

It is easy to verify that the classes of $k$-prefix codes and $k$-suffix codes, $k$-infix codes and $k$-outfix codes, are not comparable, that is, $P_k(X) \not\subseteq S_k(X)$ and $S_k(X) \not\subseteq P_k(X)$, $I_k(X) \not\subseteq O_k(X)$ and $O_k(X) \not\subseteq I_k(X)$. In fact, let $C_1 = \{a_1 a_2 \ldots a_{2k-1} a_{2k}, a_1 a_2 \ldots a_{2k}^2\}$, $C_2 = \{a_1 \ldots a_{2k}, a_1^2 \ldots a_{2k}^2\}$, where $a_i \in X$, $i = 1, \ldots, 2k$, $a_i \neq a_{i+1}$, $j = 1, \ldots, 2k - 1$. It is easy to show that $C_1 \not\subseteq P_k(X)$ and $C_1 \not\subseteq S_k(X)$, $C_2 \not\subseteq O_k(X)$ and $C_2 \not\subseteq I_k(X)$. Similarly, let $C_3 = \{a_1 \ldots a_{2k+1}, a_1 a_2^2 \ldots a_{2k+1}^2\}$, $C_4 = \{a_1 \ldots a_{2k}, a_1^2 \ldots a_{2k+1}^2\}$, where $a_i \in X$, $a_i \neq a_{i+1}$, $i = 1, \ldots, 2k + 1$, $j = 1, \ldots, 2k$. One proves that $C_3 \not\subseteq O_k(X)$ and $C_3 \not\subseteq I_k(X)$. $C_4 \not\subseteq O_k(X)$ and $C_4 \not\subseteq I_k(X)$.

Remark 2. $RE(X) \subseteq O(X)$ was not shown in Figure 1 in [2]. By Definitions 1 and 2, it is easy to see that $RE(X) \subseteq O(X)$.