F-Codes

Chun-Hua Cao, Feng-Xia Zhang
School of Mathematics and Statistics
Yunnan University, Kunming, China

Di Yang† and Qiu-Jun Liao
School of Information
Yunnan University of Finance and Economics, Kunming, China
†e-mail: yangdi65@aliyun.com

By adding special conditions, Jean Berstel in 2012 proposed some new codes, for example: F- prefix codes, F- maximal prefix codes, F- maximal codes etc.. In this paper, we discuss the properties of F- completeness, F- denseness of those codes. At first, we present some sufficient and necessary conditions for F- maximal prefix codes. Then, we show if F- dense and F- thin sets are closed under the operations of union and the production of two languages.

Keywords- F- prefix codes; F- maximal prefix codes; F- dense; F- complete.

1. Introduction

Various kinds of algebraic codes such as prefix codes, comma-free codes, codes with finite deciphering delay etc. are playing very important role in many fields of science, for example in computer science, biology, and mathematics. These codes with specific algebraic properties have been motivated and defined for different purposes in theory and applications. The study of prefix codes has researched saturation point, because many scholars from different angles and different aspects to start studying special prefix codes, for example, infix code, outfix code, maximal prefix code and so on. In [2], F- prefix code, and F- maximal prefix code are given. Since they are new codes, in this paper, we are interested in investigating the algebraic properties of F- prefix code and F- maximal prefix code.

In section three, we will show F- maximal prefix codes through F- completeness and the F- denseness. Then we construct some F- dense sets and F- thin sets.

2. Preliminaries

At first, we introduce some basic definitions which will be used later. Let A be a finite set of symbols, which is called an alphabet. An element $a \in A$ is called a letter. Any finite sequence of letters in A is called a word. We often denote a word by $x = a_1 a_2 \cdots a_n$ where $a_i \in A$ for any $i = 1, 2, \cdots n.$ Let

A^* be the set of all words. We can equip A^* with an associative operation called the concatenation of two sequences. If $w = a_1a_2\cdots a_n$ and $v = b_1b_2\cdots b_m$ are two words, then the product of them is the word $z = wv = a_1a_2\ldots a_nb_1b_2\cdots b_m$. The empty sequence is called the empty word, and denoted by 1. It is the word containing no any letter and it is the neutral element for the concatenation. So the set A^* is a free monoid generated by A. Let $x = a_1a_2\cdots a_n$, where $a_i \in A$ for $i = 1,2\cdots n$ be a word. The number of letters occur in x is called the length of the word x, which is denoted by $|x|$. We let $|1| = 0$. Then $|x| = n$ for the former word $x = a_1a_2\cdots a_n$.

For any $u, v, w \in A^*$, u is called a prefix of v if $v = uxw$ for some $x \in A^*$, which is denoted by $u \le_p v$, and w is called a suffix of v, denoted by $w \le_s v$ Let $A^+ = A^* \setminus \{1\}$. A nonempty set $X \subset A^+$ or the set $\{1\}$ is called a language. A language X is called a prefix code (or a suffix code) if $A \cap AX^+ = \varnothing \left(or\ A \cap X^+A = \varnothing\right)$.

Let X be a set of some words or let $X \subseteq A^*$. We call a word x is a prefix (or suffix) of X if x is a prefix (or suffix) of some word in X. The set X is called a prefix-closed (or suffix-closed) set if the prefixes (or suffixes) of all its words are in X. The set X is called a factorial set if the factors of all its words are in X. The set X is called a recurrent set if X is a factorial set and for all $u, w \in X$, there exists a word $v \in X$ such that $uvw \in X$. Let x be a word, we call the right (or left) order of x with respect to the set X is the number of letters such that $xa \in X$ $\left(or\ ax \in X\right)$. The set X is called a right essential set if X is a prefix-closed set and every $x \in X$ has right order at least 1. Then we know if X is a right essential set, then for any $x \in X$ and any integer $n \ge 1$, then there exists a word v of length n such that $xv \in X$. The set X is called a left essential set if X is a suffix-closed set and for all $x \in X$, x has left order at least 1.

Let $F \subseteq A^*$ and $X \subseteq F$. The set X is called right dense in F, or simply right $F-$dense, if for every word in $u \in F$, u is a prefix of X. The set X is called right complete in F, or simply right $F-$complete, if X^* is right dense in F. The set X is called thin in F, or simply F- thin, if there exists a word $u \in F$ such that u is not factorial of X. Let

$P \subseteq F$ be a prefix code. The prefix code P is called maximal in F, or simply $F-$maximal, if there exist a prefix code $Y \subseteq F$, then $Y = P$.

The notions mentioned above and other notions which did not mentioned here can find in [1,2,16,18]. In the following, we cite some results proved in [1,2].

Lemma 2.1.[2] Let $F \subseteq A^*$ be a set and $X \subseteq F$ be a prefix code. Then the following statements are equivalent.

(i) For every word in $u \in F$, there exists a word in $w \in X$ such that u is a prefix-comparable with w;

(ii) The prefix code X is $F-$maximal.

Lemma 2.2.[2] Let $F \subseteq A^*$ be a factorial set and $X \subset F$ be a nonempty set. Then the following statements are equivalent.

(i) For every word in $u \in F$, there exists a word in $w \in X$ such that u is a prefix-comparable with w;

(ii) XA^* is right $F-$dense;

(iii) X is right $F-$complete.

Lemma 2.3.[2] Let $F \subseteq A^*$ be a recurrent set and $X \subseteq F$ be $F-$thin. Then the following statements are equivalent.

(i) X is an $F-$maximal prefix and suffix (or simply bifix code) code;

(ii) X is a left $F-$complete prefix code;

(ii') X is a right $F-$complete suffix code;

(iii) X is an $F-$maximal prefix code and an $F-$maximal suffix code.

Lemma 2.4.[1] A maximal code is complete.

Lemma 2.5.[1] Let $w \in A^*$ be a word and $X \subseteq A^+$ be a maximal code. Then $X^* wA^* \bigcap X^* \neq \varnothing$.

Lemma 2.6.[1] Let M be a monoid and $P, Q, R \subseteq M$. If $P \bigcup Q$ is thin, then P and Q are all thin. If R is dense, and P is thin, then $R \setminus P$ is dense.

Lemma 2.7.[1] A thin and complete code is maximal.

Lemma 2.8.[1] Let $X \subseteq A^+$ be a code. If X is complete, then X is dense or maximal.

Lemma 2.9.[1] If X, Y are prefix codes, then XY is a prefix code. It is also hold for maximal prefix code.

Lemma 2.10.[1] Let $X \subseteq A^*$. If $X = Y \setminus YA^+$ is a prefix code, then

$XA^* = YA^*$.

3. F- Codes

Proposition 3.1. Let $F \subseteq A^*$ and $X \subseteq F$ be a code. If the following conditions hold:

(i) For every word in $u \in F$*, there exists a word in* $x \in X$ *such that* u *is a prefix-comparable with* x*;*

(ii) XA^* *is right* F*- dense;*

(iii) X *is an* F*- maximal prefix code;*

then we have the following results:

(i) and (ii) are equivalent;

(2) If X *is a prefix code, then (i), (ii) and (iii) are equivalent.*

Proof. (1) (i) $\Rightarrow$ *(ii) Since* $u \in F$ *such that* u *can be prefix-comparable with some word* $x \in X$*. So there exist* $v, w \in A^*$ *such that* $uv = xw \in XA^*$*. Then* u *is a prefix of* XA^**. Therefore, by the definition of right* F*- dense, we know* XA^* *is right* F*- dense.*

(ii) $\Rightarrow$ *(i) Since* XA^* *is right* F*- dense, then for every word* $u \in F$*, we have* u *is a prefix of* XA^**. So there exist* $v, w \in A^*$ *and* $x \in X$ *such that* $uv = xw$*. Thus* u *can be prefix-comparable with the word* $x \in X$*. That is to say, every word in* F *is prefix-comparable with some word in* X*.*

(2) If X *is a prefix code, by lemma 2.1, we know (i) and (iii) are equivalent. So the condition (i), (ii) and (iii) are equivalent.*

Theorem 3.2. Let $F \subseteq A^*$ *be a suffix-closed set, and* $X \subseteq F$ *be a code. Then the following statements are equivalent.*

(i) XA^* *is a right* F*- dense set;*

(ii) X *is a right* F*- complete set.*

Proof. (i) $\Rightarrow$ *(ii) For any word* $u \in F$*, we want to prove that* u *is a prefix of* X^**. Since* XA^* *is right* F*- dense, by lemma 3.1, we have every word in* F *is prefix-comparable with a word in* X*. Then there exist* w*,* $w' \in A^*$ *and* $x \in X$ *such that* $uw = xw'$*. If* u *is a prefix of* x*, then* u *is a prefix of* X^**. Otherwise,* x *is a prefix of* u*. So there exists* $u' \in A^*$ *such that* $u = xu'$*. Since* $u \in F$ *and* F *is a suffix-closed set, then* $u' \in F$*,*

because $x \neq 1$. Then $|u'| < |u|$. Similarly, we have XA^* is right F-dense by proposition 4.1. So every word in F is prefix-comparable with a word in X. Therefore, there exist $w_1, w_1' \in A^*$, and $x_1 \in X$ such that $u'w_1 = x_1 w_1'$. If u' is a prefix of x_1, then u' is a prefix of X^*. Otherwise, x_1 is a prefix of u'. Then there exists $u'' = A^*$ such that $u' = x_1 u''$. Since F is suffix-closed, then $u'' \in A^*$. Since X has nonempty word, then $x' \neq 1$. Thus $|u''| < |u'|$. By induction, we know the word u' is a prefix of X^*. So u is a prefix of X^*. Hence X is right $F-$complete.

(ii) $\Rightarrow$(i) Since X is right F-complete, then for any word $u \in F$, we have u is a prefix of X^*, because $X^* \subset XA^*$. Then u is a prefix of XA^*. By the definition of right F-dense, we know XA^* is right F-dense.

Proposition 3.3. Let $F \subseteq A^*$ be a set and $X \subseteq F$ be a code. Then the following statements are equivalent.

(i) X is an F-maximal prefix code;

(ii) XA^* is right F-dense.

Proof.(i)$\Rightarrow$(ii) Since X is a F-maximal prefix code, then for any word $u \in F \setminus X$, u is prefix-comparable with some prefix of some word of X. Otherwise, $X \cup \{u\}$ is a prefix code, and $X \cup \{u\} \subset F$. This contradicts that X is a $F-$maximal prefix code. So u is prefix-comparable with some prefix of some word of X. If there exists $w \in X$, such that u is prefix-comparable with w, then we consider the following two conditions. If u is a prefix of w, then u is a prefix of XA^*. This is a contradiction. If w is a prefix of u, there exists $w' \in A^*$, such that $u = ww' \in XA^*$. So u is a prefix of XA^*. From all above, we know u is a prefix of XA^*. Thus for any word $u \in F$, u is a prefix of XA^*, we know XA^* is right F-dense.

(ii)$\Rightarrow$(i) Since XA^* is right F-dense, then for any $u \in F \setminus X$, u is a prefix of XA^*. That is to say, there exist $w, w' \in A^*$, and $x \in X$, such that $uw = xw'$. So u is prefix-comparable with some prefix of some word of X. Therefore $X \cup \{u\}$ is not a prefix code, thus X is a F-maximal prefix code.

From the above proposition, we have the following corollary.

Corollary 3.4. Let $F \subseteq A^*$ *be a suffix-closed set, and* $X \subseteq F$ *be a prefix code with nonempty words. Then the following statements are equivalent.*

(i) Every element of F *is prefix-comparable with some word of* X *;*

(ii) XA^* *is right* F*- dense;*

(iii) X *is an* F*- maximal prefix code;*

(iv) X *is right* F*- complete.*

Proposition 3.5. Let $F \subseteq A^*$ *be a suffix-closed set,* $L \subseteq F$ *which contains nonempty words, and* $X = L \setminus LA^+$*. Then the following statements are equivalent.*

(i) L *is right* F*- complete;*

(ii) X *is an* F*- maximal prefix code.*

Proof. (i)$\Rightarrow$*(ii) Since* $F \subseteq A^*$ *is suffix-closed and* $L \subset F$*, and* L *has nonempty word, then we know* L *is right* F*- complete.*

$\Leftrightarrow$ *Since* LA^* *is right* F*- dense, then* $XA^* = LA^*$*. So* XA^* *is right* F*- dense. Therefore* X *is an* F*- maximal prefix code.*

(ii)$\Rightarrow$ *(i) Since* X *is an* F*- maximal prefix code, then we know* XA^* *is right* F*- dense. Then we have* $XA^* = LA^*$*. So* LA^* *is right* F*- dense. By proposition 3.2, we know* L *is right* F*- complete.*

Corollary 3.6. Let $F \subseteq A^*$ *be a suffix-closed set, and* $X \subseteq F$ *be a prefix code. Then the following statements are equivalent.*

(i) X *is right* F*- complete;*

(ii) X *is an* F*- maximal prefix code.*

Proof. Let $L = X$*. Since* X *is a prefix code, then* $X = X / XA^+$*. So it satisfies the conditions in proposition 3.5. Thus (i) is equivalent with (ii).*

Proposition 3.7. Let A *contain more than two letters and* $F \subseteq A^*$ *be a set. If* $\left\{ uab^{|u|} \middle| u \in A^+ \right\} \subseteq F$*, and* F *is a suffix-closed set, then for all* $u \in F$*, there exists* $v \in F$*, such that* $uv \in F$ *is an unbordered word.*

Proof. Assume a *be the first letter in* u*, and* $b = A \setminus \{a\}$*. If* $w = uab^{|u|}$ *is an unbordered word, and* t *is a nonempty prefix of* w *which begins as the letter* a *and* a *is not the suffix of* w*. Otherwise* $|t| = |u|$*,*

but $t = sab^{|u|}$, which $s \in A^*$, and $t = uab^{|s|}$. Therefore we have $|s| = |u|$. So $t = w$. Hence $w = uv$ is an unbordered word for $w \in F$.

From the above proposition, we have the following result.

Corollary 3.8. Let $F \subseteq A^*$ be a suffix-closed set and $\{uab^{|u|} \mid u \in A^+\} \subset F$. If $X \subseteq F$ is an F- maximal code, then $X^* wA^* \cap X^* \neq \varnothing$ for all $w \in F$.

Proposition 3.9. Let $F \subseteq A^*$ be suffix-closed and $\{uab^{|u|} \mid u \in A^+\} \subseteq F$. If $X \subseteq F$ is an F- maximal code, then X is F- complete.

Proof. Since $F \subseteq A^*$ is a suffix-closed code, then $\{uab^{|u|} \mid u \in A^+\} \subset F$, and $X \subseteq F$ is an F- maximal code. So by corollary 3.8, we know for any $w \in F$, $X^* wA^* \cap X^* \neq \varnothing$. Then for any $w \in F$, we have $A^* wA^* \cap X^* \neq \varnothing$. Thus X is F- complete.

Proposition 3.10. Let $F \subseteq A^*$ be a recurrent set, and $P, Q, R \subseteq F$. If $P \cup Q$ is F- thin, then P, Q is F- thin. If R is F- dense and P is F- thin, then $R \setminus P$ is F- dense.

Proof. $\Leftarrow)$ Since P, Q are F- thin, then there exists $m \in F$ such that $A^* mA^* \cap P = \varnothing$. So there exists $n \in F$ such that $A^* nA^* \cap Q = \varnothing$. Since F is recurrent, then there exists $w \in F$ such that $mwn \in F$. Therefore $A^* mwA^* \cap P = \varnothing$, and $A^* wnA^* \cap Q = \varnothing$. So $A^* mwnA^* \cap (P \cup Q) = \varnothing$. Hence mwn is not complete in $P \cup Q$. Hence $P \cup Q$ is F- thin.

$\Rightarrow)$ If $P \cup Q$ is F- thin, then $A^* mwnA^* \cap (P \cup Q) = \varnothing$ for some $m \in F$. So $A^* mA^* \cap P = \varnothing$, and $A^* wA^* \cap Q = \varnothing$. Therefore P, Q is F- thin. If R is F- dense, and P is F- thin, then $R \setminus P$ is not possible F- thin. Suppose $R \setminus P$ is F- thin, then $R = (R \setminus P) \cup P$ is F- thin, which contradicts R is F- dense.

Proposition 3.11. Let $F \subseteq A^*$. *Then every finite subset of* F *is* F-*thin.*

Proof. Let $X \subseteq F$ *and* X *be a finite set. Let* $m \in F$, *and* $|m| > \max\{|x| \mid x \in X\}$. *Then* $m \in F$, *and* $A^* m A^* \cap X = \varnothing$. *Hence* X *is* F-*thin.*

Proposition 3.12. Let $F \subseteq A^*$ *be a recurrent set and* $X, Y \subseteq F$ *be nonempty sets. If* X, Y *are* F-*thin, then* XY *is* F-*thin.*

Proof. Since X, Y *are* $F-$*thin, then there exists* $m \in F$ *such that* $A^* m A^* \cap X = \varnothing$. *That is to say,* $m \in \overline{F(X)}$, *there exists* $n \in F$, *and* $A^* n A^* \cap Y = \varnothing$. *Therefore* $n \in \overline{F(Y)}$, *because* $F \subset A^*$ *is recurrent. Then there exists* $w \in F$ *such that* $mwn \in F$, *because* $m \in \overline{F(X)}$. *Therefore* $mw \in \overline{F(X)}$. *Then* $mwn \in \overline{F(XY)}$. *Thus* $A^* mwn A^* \cap XY = \varnothing$. *Hence* XY *is* F-*thin.*

Proposition 3.13. Let $F \subseteq A^*$ *be a suffix-closed set, and* $X \subseteq F$ *be* F-*thin and* X *doesn't contain the empty word.*

(i) X *is an* F-*maximal code and a prefix code;*

(ii) X *is an* F-*maximal prefix code;*

(iii) X *is a right* F-*complete code.*

Then we have (i) $\Rightarrow$ *(ii)* $\Rightarrow$ *(iii).*

Proof. (i) $\Rightarrow$ *(ii) Since* X *is an* F-*maximal code and a prefix code, and* X *is* F-*thin, then* X *is an* F-*maximal prefix code.*

(ii) $\Rightarrow$ *(iii) Since* X *is an* F-*maximal prefix code, by proposition 3.3, we know* XA^* *is right* F-*dense. Because* $F \subset A^*$ *is a suffix-closed set, by proposition 3.2, we know* X *is right* F-*complete.*

4. Conclusion

After the presentation of the definitions of F-prefix codes, F-bifix codes, F-dense sets, F-thin sets and F-complete sets, we mainly discuss the equivalent conditions for F-maximal prefix codes. In the future, we are going to consider the relationship among maximal F-codes, maximal F-infix codes, maximal F-comma-free codes and maximal F-prefix codes. We

want to give some structures of finite F- maximal prefix codes.

References

[1] Jean Berstel, Dominique Perrin, Christophe Reutenauer , Codes and automata, Cambridge University Press, 2009.
[2] Jean Berstel, Bifix codes and sturmian words, Journal of Algebra, Vol. 369, 146-202, 2012.
[3] Jean Berstel, Recent results on syntactic groups of prefix codes, European Journal of Combinatorics, Vol.33, 1386-1404, 2012.
[4] Denis Derencourt, A three-word code which is not prefix-suffix composed, Theoretical Computer Science, Vol. 163, 145-160, 1996.
[5] Jean Berstel, On the groups of codes with empty kernel, Semigroup Forum, Vol.80,351-374, 2010.
[6] Isabel M.Araujo,Veronique Bruyere, Words derivated from Sturmian words, Theoretical Computer Science, Vol.340, 204-215, 2005.
[7] Aturo Carpi, Aldo de Luca, Codes of central sturmian words , Theoretical Computer Science, Vol.340, 220-239, 2005.
[8] Clelia De Felice, Finite biprefix sets of paths in a graph D, Theoretical Computer Science, Vol.58, 103-128, 1988.
[9] Andreas W.M. Dress, R. Franz, Parametrizing the subgroups of finite index in a free group and related topics, Bayreuth. Math. Schr., Vol.20(1-8), 1985.
[10] Fabien Durand, A characterization of substitutive sequences using return words, Discrete Mathematics, Vol.179 , 89-101, 1998.
[11] Amy Glen, Jacques Justin, Episturimian words: a surey, Theor. Inform. Appl., Vol.43, 403-442, 2009.
[12] Marshall Hall Jr., Subgroups of finite index in a free groups, Canad. J. Math.,Vol.1, 187-190, 1949.
[13] Jacques Justin, Giuseppe Pirillo, Episturimian words: shifts, morphisms and numeration systems, Internat. J. Found. Comput. Sci., Vol.15, 329-348, 2004.
[14] Jacques Justin, Giuseppe Pirillo, Episturimian words and episturimian morphisms, Theoretical Computer Science, Vol.276, 281-313, 2002.
[15] Jacques Justin, Laurent Vuillon, Return words in sturmian and episturimian words, Theor Inform. Appl., Vol.34, 343-356, 2000.
[16] M.Lothaire, Algebraic Combinatorics on Words, Cambridge University Press, 2000.
[17] Cao C.H., Yang D. and Liu Y., Sub-classes of the monoid of left cancellative languages,International Journal of Computer Mathematics, Vol.88, 1619-1628, 2011.
[18] Shyr H.J., Free monoids and languages, Hon Min Book Company, Taichung,

Taiwan, 2001.
[19] Bestel J. and Perrin D., Theory of codes, Academic Press, Orlando, 1985.
[20] Shyr H.J. and Tsai Y.S., Free submonoids in the monoid of languages, Discrete Mathematics, Vol.181, 213-222, 1998.
[21] Cao C.H., The problem of freedom on certain subsemigroups of the submonoid of left singular languages, International Journal of Computer Mathematics, Vol.81, 121-132, 2004.