行政院國家科學委員會專題研究計畫　成果報告

純正數碼性質之研究

計畫類別：個別型計畫
計畫編號：NSC95-2115-M-040-001-
執行期間：95 年 01 月 01 日至 95 年 07 月 31 日
執行單位：中山醫學大學資訊管理學系

計畫主持人：黃政治

報告類型：精簡報告

處理方式：本計畫可公開查詢

中　華　民　國 95 年 10 月 17 日

純正數碼性質之研究

中文摘要

對於本研究，我們除了討論純正數碼的性質之外並探究只含有兩個元素之純正數碼的特性。而後，我們發現同時具有逗點自由數碼及 d 數碼性質的數碼是固態數碼。最後，我們證明出 d 數碼、中間數碼及固態數碼都是屬於純正數碼的範疇。

關鍵詞： 逗點自由數碼、d 數碼、中間數碼、純正數碼。

Some Properties of Pure Codes[†]

C. C. Huang*

Department of Information Management Science, Chung Shan Medical University, Taichung, Taiwan 402

Abstract

In this paper, we show some properties of pure codes and give several characterizations on two-element pure codes. It can be shown that a language is a solid code if and only if it is a comma-free code which is also a d-code. We also show that d-codes, intercodes and solid codes are pure code.

Keywords: comma-free code, d-code, intercode, pure code.

1. Introductions

Codes are formal languages with special combinatorial and structural properties which are exploited in information processing or information transmission. To investigate properties of codes, one may discuss the relationships among all classes of codes. The relationships construct the hierarchy of classes of codes that we have seen from ([6], [13]). For instance, the class of the solid codes ([2], [9]) lie below the class of comma-free codes ([1]) in the hierarchy of classes of codes.

The notion of pure languages was introduced in ([5]). Subsequently, the concept of the pure code turns up the property of the preserving homomorphisms. There is compelling evidence that if the image of alphabet is a pure code by a homomorphism, then the homomorphism is primitivity-preserving-homomorphism ([8]). The concept of pure code gives rise to a motivation for studying the properties of pure codes and investigating the relationships with others. The relationship between the class of pure codes and the class of d-codes is investigated in ([3]). Therefore, beside the relationships discussed among the classes of codes, some characteristics of pure codes, comma-free codes, d-codes and solid codes will be studied in this paper.

This paper is organized into several sections. The first section introduces the overview of this paper. In the second section, we will display some well-known definitions and properties applied in this paper. To investigate the relationship among comma-free codes, d-codes and solid codes, some properties of d-codes should be explored. In the third section, we can get some characteristics of the d-codes. Moreover, we intend to explore the result that a language is a solid code if and only if it is a d-code and also a comma-free code in the fourth section. Finally, some properties of pure codes is studied. Meanwhile, as will become evidence that d-codes, solid codes, comma-free codes and intercodes are pure codes.

† This work was supported by the National Science Council R.O.C. under Grant NSC 95-2115-M-040-001. AMS Subject Classification: 68R15

* E-mail: cchuang@csmu.edu.tw

2. Definitions and Preliminaries

Let X be a finite alphabet and X^* the free monoid generated by X. Any element of X^* is called a *word*. The length of a word w is the number of letters occurring in w and denoted by $\lg(w)$. Any subset of X^* is called a *language*. Let $X^+ = X^* \setminus \{1\}$ where 1 is the empty word. A word $w \in X^+$ is said to be *primitive* if $w = f^n$ with $f \in X^+$ always implies $n = 1$. Let Q denote the set of all primitive words. For a word $w \in X^+$, there exists a unique primitive word f and a unique integer $i \geq 1$ such that $w = f^i$. Let $f = \sqrt{w}$ and call f the *root* of w. For two words $u, v \in X^+$, we denote by $v \leq_{\mathrm{p}} u$ if $v \in P(u)$ and denote by $v \leq_{\mathrm{s}} u$ if $v \in S(u)$. A word $x \in X^+$ is said to be *non-overlapping* if $x = uy = yv$ for some $y, u, v \in X^*$ implies $y = 1$. Let $D(1)$ be the set of all non-overlapping words.

For a given word $x \in X^+$, we define the following sets.

$P(x) = \{y \in X^+ | x \in yX^*\}$, $\bar{P}(x) = \{y \in X^+ | x \in yX^+\}$,
$S(x) = \{y \in X^+ | x \in X^*y\}$, $\bar{S}(x) = \{y \in X^+ | x \in X^+y\}$,
$E(x) = \{y \in X^+ | x \in X^*yX^*\}$, $\bar{E}(x) = \{y \in X^+ | x \in X^+yX^* \cup X^*yX^+\}$.

A language $L \subset X^+$ is a *code* if $x_1x_2 \cdots x_n = y_1y_2 \cdots y_m$, $x_i, y_j \in L$ implies that $m = n$ and $x_i = y_i$, $i = 1, 2, ..., n$. We review the definitions of some codes used in this paper: a code L is a *prefix code* (*suffix code*) if the condition $L \cap LX^+ = \emptyset$ ($L \cap X^+L = \emptyset$) is true. A code L is a *bifix code* if L is both a prefix code and also a suffix code. A code L is an *infix code* if for all $x, y, u \in X^*, u \in L$ and $xuy \in L$ together imply $x = y = 1$. A code L is an *intercode* if $L^{m+1} \cap X^+L^mX^+ = \emptyset, m \geq 1$. The integer m is called the index of L. An intercode of index one is called a *comma-free code*. A code L is a *d-code* if L is a bifix code and $P(L) \cap S(L) = L$. Given a set $L \subseteq X^+$, any word $w \in X^+$ can be represented as: $w = x_1y_1x_2y_2 \cdots x_ny_nx_{n+1}$, where $y_j \in L$, $j = 1, 2, \ldots, n$, $E(x_i) \cap L = \emptyset$, $i = 1, 2, \ldots, n+1$. If $E(w) \cap L = \emptyset$ or $L = \emptyset$, then we let $w = x_1$. Any such representation of w is called an L-representation of w. A code L is a *solid code* if for any $w \in X^+$ there is a unique L-representation. In the following, we review some results used in the rest of this paper.

Lemma 2.1 ([4]) Let $u, v \in Q$ with $u \neq v$. Then $u^mv^n \in Q$ for all $m \geq 2, n \geq 2$.

Lemma 2.2 ([7]) Let $uv = f^i, u, v \in X^+, f \in Q, i \geq 1$. Then $vu = g^i$ for some $g \in Q$.

Lemma 2.3 ([4]) If $uv = vu, u, v \in X^+, u \neq 1, v \neq 1$, then u, v are powers of a common word.

Lemma 2.4 ([3]) If $uv = vz, u, v, z \in X^*$ and $u \neq 1$, then $u = (pq)^i, v = (pq)^jp, z = (qp)^i$ for some $p, q \in X^*, i \geq 1, j \geq 0$ and $pq, qp \in Q$.

Lemma 2.5 ([10]) If $uq^m = g^k$ for some $m, k \geq 1, u \in X^+$ and $g \in Q$ with $u \notin q^+$. Then $q \neq g$ and $\lg(g) > \lg(q^{m-1})$.

Lemma 2.6 ([12]) Let $x_1, x_2, y_1, y_2 \in X^+$ be such that $x_1y_1 \in Q$. If $x_1y_1 = x_2y_2$ and $y_1x_1 = y_2x_2$, then $x_2 = x_1$ and $y_2 = y_1$.

Lemma 2.7 ([11]) Let X be an alphabet at least two letters and let $L \subseteq X^+$ be an intercode of index n with $n \geq 1$. Then for every m, $m \geq n$, L is an intercodes of index m.

3. Some Properties of d-codes

A language $L \subseteq X^+$ is said to be a d-code, which is introduced by Y. Y. Lin in ([3]), if L is a bifix code and $P(L) \cap S(L) = L$. To investigate the relationships among d-codes, comma-free codes and solid codes, we study some properties of d-codes in this section.

Proposition 3.1 ([3]) Let $L \subseteq X^+$. The following statements are equivalent:
(1) L is a d-code.
(2) Any proper prefix of a word in L is not a suffix of any word in L and any proper suffix of a word in L is not a prefix of any word in L.
(3) For any $u, v \in L$, if there exists $x \in X^+$ such that $x \leq_{\mathrm{p}} u, x \leq_{\mathrm{s}} v$, then $x = u = v$.

From the above result, the following conclusions are given immediately.

Proposition 3.2 If L is a d-code, then for every $u, v \in L$, $P(u) \cap S(v) \neq \emptyset$ implies $u = v$.

Proof. It is clear. $\square$

Proposition 3.3 Let $L \subseteq X^+$. L is a d-code if and only if L is a bifix code and $\bar{P}(L) \cap \bar{S}(L) = \emptyset$.

Proof. It is clear from Proposition 3.2 and the definition of d-codes. $\square$

4. The Relationships of Families of d-codes, Comma-free Codes and Solid Codes

In ([2]), Shyr had shown that a solid code is a comma-free code. The converse is not true. For example, $\{aba\}$ is a comma-free code. But the word $ababa = ab(aba) = (aba)ba$ has two different L-representations, by the definition of solid codes, $\{aba\}$ is not a solid code. Beside the known relationship which the family of solid code is contained in the family of comma-free code, the relationships among the families of d-codes, comma-free codes and solid codes will be studied in this section. In fact, the family of solid codes is the intersection of the family of comma-free codes and the family of d-codes. Before the result is explored, we review some characteristics of solid codes used in this section.

Proposition 4.1 ([1]) Let X be an alphabet and let $L \subseteq X^+$. If L is a solid code, then L is a comma-free code and hence an infix code.

Proposition 4.2 ([10])L is a solid code if and only if every two words $u, v \in L$ satisfy the following conditions:
(1) $\bar{P}(u) \cap \bar{S}(v) = \emptyset$.
(2) If $u \neq v$, then $u \notin E(v)$ and $v \notin E(u)$.

In the following, the relationships are investigated among solid codes, comma-free codes, and d-codes.

Proposition 4.3 Let $L \subseteq X^+$. Then the following statements are equivalent:
(1) L is a solid code.
(2) L is an infix d-code.
(3) L is a comma-free code with $\bar{P}(L) \cap \bar{S}(L) = \emptyset$.

Proof. ((2) $\Rightarrow$(1)) Suppose that L is an infix d-code. Since L is a d-code, by Proposition 3.2, $\bar{P}(u) \cap \bar{S}(v) = \emptyset$ for every $u, v \in L$. In the meanwhile, since L is an infix code, either $u \in E(v)$ or $v \in E(u)$ will imply $u = v$. Hence, by Proposition 4.2, L is a solid code .

((1) $\Rightarrow$ (2)) Suppose L is a solid code. Then, by Proposition 4.1, L is an infix code and hence a bifix code. Since L is an infix code, either $u \in E(v)$ or $v \in E(u)$ implies that $u = v$ for any $u, v \in L$. Thus $\bar{P}(u) \cap \bar{S}(v) = \emptyset$ for all $u, v \in L$ with $u \neq v$. Hence, by Proposition 3.3, L is a d-code.

((3) $\Rightarrow$ (1)) Since L is a comma-free code, L is an infix code. Then either $u \in E(v)$ or $v \in E(u)$ implies that $u = v$ for every $u, v \in L$. This in conjunctive with $\bar{P}(L) \cap \bar{S}(L) = \emptyset$ and Proposition 4.2 yields that L is a solid code.

((1) $\Rightarrow$ (3)) Suppose that L is a solid code. Then by Proposition 4.1 and 4.2, L is a comma-free code with $\bar{P}(L) \cap \bar{S}(L) = \emptyset$. □

Proposition 4.4 Let $L \subseteq X^+$. L is a solid code if and only if L is a d-code and also a comma-free code.

Proof. ($\Rightarrow$) From Proposition 4.1 and 4.3, the result is clear.
($\Leftarrow$) Let L be a d-code and a comma-free code. Suppose that L is not a solid code. Then, by Proposition 4.2, there exist $u \neq v \in L$ such that $\bar{P}(u) \cap \bar{S}(v) \neq \emptyset$, $v \in E(u)$ or $u \in E(v)$. Since $\bar{P}(u) \cap \bar{S}(v) \neq \emptyset$, L is not a d-code. In the meanwhile, either $u \in E(v)$ or $v \in E(u)$ implies that L is not a comma-free code. Therefore, as L is a d-code and also a comma-free code, it implies that L is a solid code. □

5. Pure Codes

A language L is *pure code* if it is a code such that for any $x \in L^*$, $\sqrt{x} \in L^*$. To investigate the relationship between pure code and others. In this section, the characteristics of pure codes will be investigated first. We have the following result.

Proposition 5.1 Every pure code is contained in Q.

Proof. Let L be a pure code. Suppose that L is singleton, that is, $L = \{u\}$. If $u = f^n$ with $n \geq 2$ where $f \in Q$, then $u \in L^*$ but $\sqrt{u} = f \notin L^*$. Hence $L = \{u\}$ is not pure, a contradiction. Thus $u \in Q$. Next, suppose that L contains two or more words. Let $u = f^i \in L, i \geq 2$. Since L is pure, $f = \sqrt{u} \in L^*$. Let $f = u_1 u_2 \cdots u_n, u_k \in L \setminus \{u\}, 1 \leq k \leq n$. Thus $u = f^i = (u_1 u_2 \cdots u_n) \cdots (u_1 u_2 \cdots u_n)$. This contradicts that L is a code. □

Let $A \subseteq X^+$ be a code. A word $x \in A^*$ is a *root word* constructed by A if $x = f^n, n \geq 1, f \in A^+$ always implies $n = 1$ and $x = f$. For a code $A \subseteq Q$, let Q_A be the set of all root words constructed by A. Similar to the definitions of $Q^{(i)}, i \geq 1$, $Q_A^{(1)} = \{1\} \cup Q_A$ and $Q_A^{(i)} = \{f^i | f \in Q_A\}$ for any $i \geq 2$. For any $x \in A^*$, it is clear that if $x \in Q$, then $x \in Q_A$. But the converse is not true. For example, let $A = \{aba, b\}$. Then the word $x = (aba)b = (ab)^2 \in A^*$. By the definition of root word constructed by A, $(ab)^2 \in Q_A$, but $(ab)^2 \notin Q$. In the following proposition, we will show that A is pure if and only if for every $x \in A^*$, $x \in Q_A$ implies $x \in Q$.

Proposition 5.2 Let X be an alphabet and $A \subseteq Q$ be a code. Then A is pure if and only if every root word x constructed by A is primitive.

Proof. Assume that A is pure. Suppose that there exists a word in Q_A which is not primitive. Let $x \in A^+$ and $x \in Q_A \setminus Q$. Since $x \notin Q$, let $x = f^n, n \geq 2, f \in Q$. However, $x \in Q_A$ implies $f \notin A^*$. That is, $x \in A^*$ and $\sqrt{x} \notin A^*$. It contradicts that A is pure. Conversely, suppose that A is not pure. Then there exists $x \in A^*$ such that $\sqrt{x} \notin A^*$. (1) If $x \in Q_A$, then $x \notin Q$. Indeed, if $x \in Q$, then $x \in Q_A$. This implies that $\sqrt{x} = x \in A^*$, a contradiction. (2) If $x \notin Q_A$, then there exist $g \in Q_A$ and $m \geq 2$ such that $x = g^m$. Since $x = f^n, f \notin A^*$, we have $g^m = f^n$. But from $g \in Q_A \subseteq A^*$ and $f \notin A^*$, there exists $i \geq 2$ such that $g = f^i$. Thus $g \in Q_A$ and $g \notin Q$, a contradiction. $\square$

Proposition 5.3 Let $L = \{u, v\} \subseteq Q$. If L is a pure code, then $uv \in Q$.

Proof. Suppose that $uv \notin Q$. Let $uv = f^n, n \geq 2, f \in Q$. Then there exist $f_1, f_2 \in X^+$ such that $u = f^k f_1, v = f_2 f^{n-k-1}$ with $f = f_1 f_2$ where $n - 1 \geq k \geq 0$. We have the following cases:

(1) $k = 0$. Then $u = f_1, v = f_2 f^{n-1}$. Thus $\lg(u) < \lg(f) < \lg(v)$. This case implies that $f \notin L^*$.

(2) $n - 2 \geq k \geq 1$. Since $u = f^k f_1, v = f_2 f^{n-k-1}$, we get $\lg(f) < \lg(u)$ and $\lg(f) < \lg(v)$. This case also implies that $f \notin L^*$.

(3) $k = n - 1$. Then $u = f^k f_1, v = f_2$. Thus $\lg(v) < \lg(f) < \lg(u)$. This case also implies that $f \notin L^*$.

From (1), (2) and (3), $uv \in L^* \setminus Q$ implies $\sqrt{uv} \notin L^*$. This contradicts that L is pure code. Hence $uv \in Q$. $\square$

In the following, we give a characterization for $\{u, v\} \subseteq X^n$ which is a pure code. It needs the following property.

Lemma 5.4 Let $x, y \in X^+$. Then $yx \leq_{\mathrm{p}} x^m y, m \geq 1$ if and only if $\sqrt{x} = \sqrt{y}$.

Proof. ($\Leftarrow$) Immediate.
($\Rightarrow$) Let $yx \leq_{\mathrm{p}} x^m y$ and $x, y \in X^+$. If $m = 1$, then $yx = xy$. Thus, by Lemma 2.3, $\sqrt{x} = \sqrt{y}$. So we will consider the case $m \geq 2$. There are the following three cases:

(1) $y \leq_{\mathrm{p}} x^{m-1}$. If $y = x^{m-1}$, then $\sqrt{x} = \sqrt{y}$. We consider $y <_{\mathrm{p}} x^{m-1}$. Note that $yx <_{\mathrm{p}} x^m$. There exists $x_1, x_2 \in X^+$ with $x = x_1 x_2$ such that $yx = x^j x_1$, where $1 \leq j \leq m - 1$. Since $yx = yx_1 x_2 = x^j x_1 = x_1 (x_2 x_1)^j$, we get $x_1 x_2 = x_2 x_1$. Hence, by Lemma 2.3, $\sqrt{x_1} = \sqrt{x_2} = \sqrt{x}$. This in conjunctive with $y = x^{j-1} x_1$ yields that $\sqrt{y} = \sqrt{x_1} = \sqrt{x}$.

(2) $x^{m-1} <_{\rm p} y \leq_{\rm p} x^m$. If $y = x^m$, then $\sqrt{x} = \sqrt{y}$. Hence $x^{m-1} <_{\rm p} y <_{\rm p} x^m$. There exist $x_1, x_2 \in X^+$ such that $y = x^{m-1}x_1$ and $x = x_1x_2$. Since $y <_{\rm p} x^m$, $yx <_{\rm p} x^{m+1}$. Thus $x^{m-1}x_1x_1x_2 <_{\rm p} x^{m-1}(x_1x_2)^2$. This yields that $x_1x_2 = x_2x_1$. Hence, by Lemma 2.3 again, $\sqrt{x_1} = \sqrt{x_2} = \sqrt{x}$. This in conjunctive with $y = x^{m-1}x_1$ yields that $\sqrt{y} = \sqrt{x_1} = \sqrt{x}$.

(3) $x^m <_{\rm p} y$. There exist $y_1, y_2 \in X^+$ such that $y = y_1y_2$ and $y_1 = x^m$. Since $yx \leq_{\rm p} x^m y$, $y_2 x \leq_{\rm p} y = x^m y_2$. If $\lg(y_2) \leq \lg(x^m)$, then this condition is similar to Cases (1) and (2). It yields that $\sqrt{x} = \sqrt{y_2}$. This in conjunctive with $y = x^m y_2$ yields that $\sqrt{y} = \sqrt{x}$. If $\lg(y_2) > \lg(x^m)$, then there exist $k \geq 1, y' \in X^+$ such that $y = (x^m)^k y'$ where $\lg(y') \leq \lg(x^m)$. Again, it is similar to Cases (1) and (2) and yields $\sqrt{x} = \sqrt{y}$. □

Proposition 5.5 Let $u \neq v \in Q$ with $\lg(u) = \lg(v)$. Then $\{u, v\}$ is pure code if and only if $uv \in Q$.

Proof. ($\Rightarrow$) By Proposition 5.3, the result is clear.

($\Leftarrow$) Let $uv \in Q$. Suppose that $L = \{u, v\}$ is not a pure code. Then there exists a word $x \in L^* \setminus Q$ with minimal length such that $x = x_1x_2 \cdots x_m = f^n$, where $f \in Q \setminus Q_L, m, n \geq 2, x_i \in \{u, v\}, 1 \leq i \leq m$. Since $\sqrt{x} = f \notin L^*$, we have that $f^+ \notin u^* \cup v^*$. This implies that $uv \in E(x)$ or $vu \in E(x)$. From Lemma 2.2, the only considered case is that $x_1 = u$ and $x_m = v$. That is, $u <_{\rm p} x = f^n, v <_{\rm s} x = f^n$. We will consider the following cases:

(1) $\lg(f) = \lg(u) = \lg(v)$. Then $u = f = v$, this contradicts to $u \neq v$.

(2) $\lg(f) < \lg(u) = \lg(v)$. By Lemma 2.2, let $x = u^{i_1}v^{j_1} \cdots u^{i_r}v^{j_r}$, where $i_k, j_k \geq 1, 1 \leq k \leq r$ with $i_1 + \cdots + i_r + j_1 + \cdots + j_r = m$, $i_1 \geq i_k$ for all k. As $i_1 \geq 2$, by Lemma 2.1, we have that $v^{j_1} \cdots u^{i_r}v^{j_r} \in Q$. And by Lemma 2.5, $\lg(f) > \lg(u^{i_1-1}) \geq \lg(u)$. This contradicts that $\lg(f) < \lg(u)$. Hence $i_1 = 1$. This implies that $i_k = 1$ for $1 \leq k \leq r$. Now by Lemma 2.2, we let $x = uv^{j_1} \cdots uv^{j_r}$, where $j_r \geq j_k$ for all k. It yields that $u^2 \notin E(x)$. As $j_r \geq 2$, by Lemma 2.1, we have that $uv^{j_1} \cdots u \in Q$. And by Lemma 2.5, $\lg(f) > \lg(v^{j_r-1}) \geq \lg(v)$. This contradicts that $\lg(f) < \lg(v)$. Hence $j_r = 1$ and $j_k = 1$ for $1 \leq k \leq r$. It yields that $v^2 \notin E(x)$. Since $u^2, v^2 \notin E(x)$ and $\lg(x)$ is minimal, $x = uv = f^n$. This contradicts that $uv \in Q$.

(3) $\lg(f) > \lg(u) = \lg(v)$. There exists a number $1 < k < m$ such that $f = x_1x_2 \cdots x_{k-1}x_{k1}$, $f^{n-1} = x_{k2}x_{k+1} \cdots x_m$ and $x_k = x_{k1}x_{k2}$ where $x_{k1}, x_{k2} \in X^+$. Since $L = \{u, v\} \subset Q$ and $\lg(u) = \lg(v)$, the case $x_k = u$ will be considered. The other case $x_k = v$ is similar. To consider the case $x_k = u$, we have the following two cases:

(3-1) $x_{k+1} = u$. Since $x_1 = u = x_{k1}x_{k2} <_{\rm p} f$ and $x_{k2}x_{k+1} \cdots x_m = f^{n-1}$, this yields that $x_{k2}x_{k1} = x_{k1}x_{k2}$. By Lemma 2.3, x_{k1}, x_{k2} are powers of a common word. Thus $u = w_{k1}w_{k2} \notin Q$. This contradicts that $u \in Q$.

(3-2) $x_{k+1} = v$. Since $\lg(u) = \lg(v)$, there exists a word $v_1 <_{\rm p} v$ such that $u = x_{k1}x_{k2} = x_{k2}v_1$. By Lemma 2.4, we have that

$$x_{k1} = (pq)^{i_1}, x_{k2} = (pq)^{j_1}p, v_1 = (qp)^{i_1} \qquad (4-2)$$

for some $p, q \in X^*$ with $pq \in Q$ and $i_1 \geq 1, j_1 \geq 0$. Thus $u = (pq)^{i_1+j_1+1}p$. We consider two subcases:

(3-2-1) $x_{k-1} = u$. Since $x_m = v <_{\mathrm{s}} f$ and $f = x_1x_2\cdots x_{k-1}x_{k1}$, there exists a word $u_2 <_{\mathrm{s}} x_{k-1} = u$ such that $v = u_2x_{k1}$. This in conjunctive with $\lg(u) = \lg(v)$ yields that $\lg(x_{k2}) = \lg(u_2)$ and $u_2 = x_{k2}$. Thus $v = x_{k2}x_{k1}$. From Equation (4-2), $v = (pq)^{j_1}p(pq)^{i_1}$. Since $v_1 <_{\mathrm{p}} v$, $(qp)^{i_1} <_{\mathrm{p}} (pq)^{j_1}p(pq)^{i_1}$. If $j_1 \geq 1$, then $pq = qp$. By Lemma 2.3, u, v are powers of a common word. This implies that $uv \notin Q$. This contradicts that $uv \in Q$. If $j_1 = 0$, then $u = (pq)^{i_1+1}p$. Since $(qp)^{i_1} <_{\mathrm{p}} p(pq)^{i_1}$ and $i_1 \geq 1$, we get $qp <_{\mathrm{p}} ppq$. By Lemma 5.4, we have that $\sqrt{p} = \sqrt{q}$, i.e., u, v are powers of a common word. Again, this implies that $uv \notin Q$ and this contradicts that $uv \in Q$.

(3-2-2) $x_{k-1} = v$. Since $x_m = v <_{\mathrm{s}} f = x_1x_2\cdots x_{k-1}x_{k1}$ and $\lg(u) = \lg(v)$, there exists a word $v_1, v_2 \in X^+$ with $v = v_1v_2$ such that $v = v_2x_{k1}$. That is, $v_1v_2 = v_2x_{k1}$. By Lemma 2.4,

$$v_2 = (rs)^{j_2}r, v_1 = (rs)^{i_2}, x_{k1} = (sr)^{i_2} \qquad (4-3)$$

where $r, s \in X^*$ with $rs \in Q$ and $i_2 \geq 1, j_2 \geq 0$. From Equations (4-2) and (4-3), $(qp)^{i_1} = v_1 = (rs)^{i_2}$ and $(pq)^{i_1} = x_{k1} = (sr)^{i_2}$. Since $pq, qp, rs, sr \in Q$, $qp = \sqrt{v_1} = rs$ and $pq = \sqrt{x_{k1}} = sr$. By Lemma 2.6, $r = q$ and $s = p$. Thus $u = (pq)^{i_1+j_1+1}p$ and $v = (qp)^{i_2+j_2+1}q$. This implies that $uv = (pq)^{i_1+j_1+i_2+j_2+3} \notin Q$, a contradiction. □

In the following, we will show that d-codes, solid codes, comma-free codes and intercodes are pure codes. Before we explore the relationships between pure code and others, it can be shown that every subset of a pure code is a pure code and the intersection of two pure codes is a pure code.

Lemma 5.6 Let A be a pure code and $B \subseteq A$. If $B \neq \emptyset$, then B is a pure code.

Proof. Let $w \in B^*$. If $w = 1$, then $\sqrt{w} = 1 \in B^*$. Let $w \in B^+$. It can get that $w \in B^+ \subseteq A^+$. Suppose that $\sqrt{w} \notin B^*$. Since A is a pure code, by the definition of the pure code, $\sqrt{w} \in A^*$. Let $\sqrt{w} = w_1w_2\cdots w_i\cdots w_n$ where $w_i \in A$ for some i . Then there exists code word $w_k \in A$ occurred in $\sqrt{w}$ such that w_k occurred in w for some k. This contradicts that $w \in B^+$. Hence $\sqrt{w} \in B^*$. We can conclude that B is a pure code. □

Lemma 5.7 Let A, B be pure codes and $C = A \cap B$. If $C \neq \emptyset$, then C is a pure code.

Proof. Let $w \in C^*$. Since $w \in C^+ = (A \cap B)^+ \subseteq A^+$ and A is a pure code, by Lemma 5.7, one have that $\sqrt{w} \in A^+$. Similarly, $\sqrt{w} \in B^+$. That is, $\sqrt{w} \in A^+ \cap B^+$. This implies that $\sqrt{w} \in (A \cap B)^+ = C^+$. Hence C is a pure code. □

Proposition 5.8 ([3]) Every d-code is a pure code.

Proposition 5.9 An intercode of index greater than or equal to 2 is a pure code.

Proof. Let L be an intercode of index $m \geq 2$. Suppose that L is not a pure code. Then there exists a word $w \in L^*$ such that $\sqrt{w} \notin L^*$. This implies that $\sqrt{w} \neq w$. Thus

$w \notin Q$. Suppose that $w = f^i, i \geq 2$ where $f \in Q$. Note that $\sqrt{w} = f \notin L^*$. Let $w = u_1 u_2 \cdots u_n = f^i$, where $n \geq 1$ and $u_j \in L$ for $1 \leq j \leq n$ and for every $j < n$,

$$u_1 u_2 \cdots u_j \notin f^+. \qquad (4-4)$$

We consider the following three cases: (1) $n = 1$; (2) $n = 2$; (3) $n \geq 3$.

(1) $n = 1$. Then $w = u_1 = f^i, i \geq 2$. This contradicts that L is an intercode.

(2) $n = 2$. Then $w = u_1 u_2 = f^i$, $i \geq 2$. Since L is an intercode and $f \notin L^*$, there exist $f_1, f_2 \in X^+, i_0, j_0 \geq 0$ such that $u_1 = f^{i_0} f_1, u_2 = f_2 f^{j_0}$ where $f = f_1 f_2$ and $i_0 + j_0 = i - 1$. Let $m \leq 2k$ where $k \geq 1$.

(2-1) If $i_0 = 0$, then $u_1 = f_1, u_2 = f_2 f^{j_0}, j_0 \geq 1$ and $u_1 u_2 = f^{j_0+1}$. As we consider $(u_2 u_1)^k u_2 \in L^{2k+1}$,

$$\begin{aligned}(u_2 u_1)^k u_2 &= (f_2 f^{j_0} f_1)^k f_2 f^{j_0} = f_2 (f^{j_0} f_1 f_2)^k f^{j_0} = f_2 (f^{j_0+1})^k f^{j_0} \\ &= f_2 (u_1 u_2)^k f^{j_0} \in f_2 L^{2k} f^{j_0} \subseteq X^+ L^{2k} X^+.\end{aligned}$$

Thus L is not an intercode of index of $2k$. Since $m \leq 2k$, by Lemma 2.7, L is not an intercode of index m, a contradiction.

(2-2) If $i_0 \geq 1$, then $u_1 = f^{i_0} f_1, u_2 = f_2 f^{j_0}, j_0 \geq 0$ and $u_1 u_2 = f^{i_0+j_0+1}$. As we consider $(u_1 u_2)^k u_1 \in L^{2k+1}$,

$$\begin{aligned}(u_1 u_2)^k u_1 &= (f^{i_0} f_1 f_2 f^{j_0})^k f^{i_0} f_1 = (f^{i_0+j_0+1})^k f^{i_0} f_1 = f (f^{i_0+j_0+1})^k f^{i_0-1} f_1 \\ &= f (u_1 u_2)^k f^{i_0-1} f_1 \in f L^{2k} f^{i_0-1} f_1 \subseteq X^+ L^{2k} X^+.\end{aligned}$$

Thus L is not an intercode of index of $2k$. Since $m \leq 2k$, by Lemma 2.7, L is not an intercode of index m, a contradiction.

(3) $n \geq 3$. From Equation (4-4), for any $j = 2, 3, \ldots, n-1$, $u_j \not\leq_{\mathrm{p}} f$, $u_j \not\leq_{\mathrm{s}} f$ and $f \not\leq_{\mathrm{p}} u_j$, $f \not\leq_{\mathrm{s}} u_j$. Hence for any $j = 2, 3, \ldots, n-1$, there exist $f_1, f_2, f_3, f_4 \in X^+$ and $r, s, t \geq 0$ such that $f = f_1 f_2 = f_3 f_4$ and $u_1 u_2 \cdots u_{j-1} = f^r f_1, u_j = f_2 f^t f_3, u_{j+1} \cdots u_n = f_4 f^s$. Thus $u_1 u_2 \cdots u_n = f^{r+s+t+2}$. As we consider $(u_j u_{j+1} \cdots u_n u_1 \cdots u_{j-1})^k u_j \in L^{nk+1}$,

$$\begin{aligned}(u_j u_{j+1} \cdots u_n u_1 \cdots u_{j-1})^k u_j &= (f_2 f^t f_3 f_4 f^s f^r f_1)^k f_2 f^t f_3 \\ &= (f_2 f^{r+s+t+1} f_1)^k f_2 f^t f_3 \\ &= f_2 (f^{r+s+t+1} f_1 f_2)^k f^t f_3 \\ &= f_2 (f^{r+s+t+2})^k f^t f_3 \\ &\in f_2 L^{nk} f^t f_3 \\ &\subseteq X^+ L^{nk} X^+.\end{aligned}$$

Thus L is not an intercode of index of nk. Since $m \leq 2k < nk$, by Lemma 2.7, L is not an intercode of index m, a contradiction.

From above cases (1), (2) and (3), we can conclude that L is a pure code. □

An intercode of index one is called a comma-free code. From Lemma 2.7, comma-free code is the subset of the intercode with index greater than 2. This conjunctive with Lemma 5.6 and Proposition 5.9 yields the following result.

Corollary 5.10 Every comma-free code is a pure code.

Moreover, from Proposition 4.4, a solid code is the intersection of a d-code and a comma-free code. This conjunctive with Lemma 5.7 yields the following result.

Corollary 5.11 Every solid code is a pure code.

References

[1] C. Y. Hsieh, S. C. Hsu and H. J. Shyr, Some Algebraic Properties of Comma-Free Codes, *RIMS Kenkyuroku*, Vol.**697**, Japan, (1989), 57–66.

[2] H. Jürgensen and S. S. Yu, Solid Codes, *J. Inf. Process. Cybern.*, **EIK 26**, No.**10** (1990), 563–574.

[3] Y. Y. Lin, Properties of Words and the Related Homomorphisms, M.S. Thesis, Institute of Applied Mathematics, Chung-Yuan University (1995).

[4] R. C. Lyndon and M. P. Schützenberger, On the Equation $a^M = b^N c^P$ in a Free Group, *Michigan Math. J.*, Vol.**9** (1962), 289–298.

[5] A. Restivo, On a Question of McNaughton and Papert, *Information and Control*, Vol.**25**, (1974), 93–101.

[6] G. Rozenberg and A. Salomaa (Eds.), Handbook of Formal Languages, Vol. **1**, *Word, Language, Grammar*, Springer-Verlag, (1997).

[7] H. J. Shyr and G. Thierrin, Disjunctive Languages and Codes, Fundamentals of Computation Theory, Proceedings of the 1977 Inter. FCT-Conference, Poznan, Poland, Lecture Notes in Computer Science, No.**5**, Springer-Verlag, (1977), 171–176.

[8] H. J. Shyr and G. Thierrin, Codes, Languages and MOL Schemes, *R.A.I.R.O. Informatique Theorique/Theoretical Computer Science*, Vol.**1**, No.**4**, (1977), 293–301.

[9] H. J. Shyr and S. S. Yu, Solid Codes and Disjunctive Domains, *Semigroup Forum*, Vol.**41**, No.**1** (1990), 23–37.

[10] H. J. Shyr and S. S. Yu, Non-prinmitive Words in the Language p^+q^+, *Soochow Journal of Mathematics*, Vol.**20**, No.**4** (1994), 535–546.

[11] S. S. Yu, A Characterization of Intercodes, *Intern. J. Comput. Math.*, Vol.**36** (1990), 39–48.

[12] S. S. Yu, Palindrome Words and Reverse Closed Languages, *Intern. J. Comput. Math.*, Vol.**75** (2000), 389–402.

[13] S. S. Yu, *Languages and Codes*, Tsang Hai Book Publishing Co., Taichung, Taiwan, (2005).