Appl. Math. J. Chinese Univ.
2024, 39(4): 596-603

A free monoid containing all prefix codes

CAO Chun-hua[1] LIU Qin[1,2] YANG Di[3]

Abstract. In this paper, we exhibit a free monoid containing all prefix codes in connection with the sets of i-th powers of primitive words for all $i \geq 2$. This extends two results given by Shyr and Tsai in 1998 at the same time.

§1 Introduction

Prefix codes are widely used in information theory and computer science, for example, in encoding and decoding, data compression and transmission, DES and Huffman's algorithms (see [1-4]). So they are especially hot topics in theoretical researches and practical applications. There are several equivalent ways to define a free monoid and we suggest adopting the following one: a monoid S is free if and only if every element s in S has a unique factorization as a product of elements of $\mathring{S} = S\backslash\{1\})\backslash S\backslash\{1\})^2$, the so-called basis of S. Let X be an alphabet and M be the monoid of languages over X which is not free. But the family of all prefix codes over X denoted by $P(X)$ is a free submonoid of M (see Proposition 2.17 of [5]). From then on, people are devoted to establishing more and more smaller or larger free submonoids of M than it. In [6], some submonoids of $P(X)$ are proposed. In [7], the submonoid generated by the basis of $P(X)$, denoted by $P(\mathring{X})$, together with the set of all primitive words over X, denoted by Q, is proved to be not free. In [8], the submonoid generated by $P(\mathring{X}) \cup \{Q^{(i)}\}$ is free for arbitrary $i \geq 2$, where $Q^{(i)} = \{f^i \mid f \in Q\}$ is the set of i-th powers of all primitive words. Also in [8], the submonoid generated by $P_f(\mathring{X}) \cup \{Q^{(i)} \mid i = 2, 3, 4, \ldots\}$ is free, where $P_f(\mathring{X})$ is the family of all irreducible finite prefix codes. This leads to the following natural question. Is the submonoid generated by $P(\mathring{X}) \cup \{Q^{(i)} \mid i = 2, 3, 4, \ldots\}$ free? Here in this paper, we solve the problem by proving that $P(\mathring{X}) \cup \{Q^{(i)} \mid i = 2, 3, 4, \ldots\}$ is a code, so the monoid generated by $P(\mathring{X}) \cup \{Q^{(i)} \mid i = 2, 3, 4, \ldots\}$ is a free submonoid of M. It is a free submonoid of M containing all prefix codes in connection with the sets of ith powers of primitive words for all $i \geq 2$.

Received: 2020-09-30. Revised: 2022-09-19.
MR Subject Classification: 20M35, 68Q70, 94A45.
Keywords: prefix code, free monoid, primitive word, maximal prefix code.
Digital Object Identifier(DOI): https://doi.org/10.1007/s11766-024-4279-1.
Supported by the National Natural Science Foundation of China(11861071).

§2 Preliminaries

Let X be a nonempty finite set of letters, named by an alphabet, and $|X|$ be the number of letters in X. Any finite string over X is called a word. For example, $w = a^3bab^2a$ is a word over the alphabet $X = \{a, b\}$. The word which contains no letter is called the empty word, noted by 1. The set of all words over X is noted by X^*. For any $w_1,\ w_2 \in X^*$, let $w_1 = a_1a_2\dots a_n,\ w_2 = b_1b_2\dots b_m$ where $a_i,\ b_j \in X$ for $i = 1, 2, \dots, n,\ j = 1, 2, \dots, m$, we define the catenation operation of w_1 and w_2 is the word $w_1w_2 = a_1a_2\dots a_nb_1b_2\dots b_m$. Then X^* is a monoid with the catenation. For any word $w \in X^*$, let $lg(w)$ be the number of letters occurring in w, and $lg(1) = 0$. Then $lg(w) = 8$ for the former word $w = a^3bab^2a$. Let $X^+ = X^* \setminus \{1\}$. Any nonempty subset of X^+ is a language and $\{1\}$ is a language. Let $M = \{A \mid A \subseteq X^+ \text{ or } A = \{1\}\}$. For any $A,\ B \in M$, the catenation operation of A and B is the language $AB = \{xy \mid x \in A,\ y \in B\}$. Then M is a monoid with the catenation, called the monoid of languages. For any $A \in M$, let $\underline{A} = \{x \in A \mid lg(x) \le lg(y) \text{ for all } y \in A\}$, which is the set of all words with minimal length in A. Let $lg(A) = lg(x)$ for any $x \in \underline{A}$, then $lg(AB) = lg(A) + lg(B)$ for any $A, B \in M$. So M is a monoid with length. A nonempty family of languages $\alpha \subseteq M$ is called a code if $A_1A_2\cdots A_n = B_1B_2\cdots B_m$ and $A_i,\ B_j \in \alpha$ for $i = 1, 2, \dots, n,\ j = 1, 2, \dots, m$ implies that $n = m$ and $A_i = B_i$ for $i = 1, 2, \dots, n$. If α is a code, then α generates a free monoid α^* of M. If S is a free monoid, then $\mathring{S} = (S\backslash\{1\})\backslash(S\backslash\{1\})^2$ is a code.

For any $u,\ v \in X^*$, if there exists a word $x \in X^*$ such that $ux = v$ (or $xu = v$), then u is called a prefix (or suffix) of v, denoted by $u \le_p v$ (or $u \le_s v$). We write $u <_p v$ (or $u <_s v$) if $u \le_p v$ (or $u \le_s v$) but $u \ne v$ and $u \ne 1$. A language A is called a prefix code if $A \cap AX^+ = \emptyset$, that is for any two distinct words $x,\ y$ in A, x is not a prefix of y and y is not a prefix of x. A prefix code A is called a maximal prefix code if and only if for any $x \in X^+ \setminus A$, $A \cup \{x\}$ is not a prefix code. Let $w \in X^+$, if $w = f^n$ where $f \in X^+$ and $n \ge 1$ is an integer implying that $n = 1$ and $w = f$, then w is called a primitive word. Let Q be the set of all primitive words over X and $Q^{(i)} = \{f^i \mid f \in Q\}$ for each $i \ge 2$. A word which is not a primitive word is called an imprimitive word. Each imprimitive word is in a unique $Q^{(i)}$ for some $i \ge 2$.

Some definitions which are used in the paper but not stated here can be found in [5, 9, 10].

§3 Main Result

If $|X| = 1$, let $X = \{a\}$, then $P(X) = \{\{a^i\} \mid i = 1, 2, 3, \dots\}$ and $Q = \{a\}$, $Q^{(i)} = \{a^i\}$ for each $i \ge 2$. So $P(\mathring{X}) \cup \{Q^{(i)} \mid i = 2,\ 3,\ 4,\ \dots\} = P(X)$. In the rest of the paper, we always let X be an alphabet containing at least 2 letters. For two distinct letters $a,\ b \in X$, since $a^i,\ (a^ib)^i \in Q^{(i)}$, then $Q^{(i)}$ is not a prefix code for each $i \ge 1$. So it is an interesting thing to check whether the submonoid generated by all prefix codes in connection with the sets of $Q^{(i)}$ for all $i \ge 2$ is free or not? In order to answer this question, at first, we propose some properties on prefix codes and $Q^{(i)}$, which will be used in the proof of the main result in the section.

Lemma 3.1. (*Proposition 2.11 of* [5]) *Let A_i be nonempty languages over X for $i = 1, 2, 3, \ldots, n$. If $A_1A_2 \ldots A_n$ is a prefix code, then $A_2A_3 \ldots A_n$, $A_3A_4 \ldots A_n$, $\ldots$, $A_{n-1}A_n$, A_n are prefix codes.*

Lemma 3.2. (*Proposition 1 of* [11]) *For any $u \in X^+$ and $a,\ b \in X$ where $|X| \geq 2$ and $a \neq b$, at least one of the words ua and ub is primitive.*

Proposition 3.3. *Let $A,\ B \in M \setminus \{1\}$. Then for every $i \geq 1$, none of the sets $AQ^{(i)}$, $Q^{(i)}A$, $AQ^{(i)}B$ is a prefix code.*

Proof. Assume $AQ^{(i)}$ is a prefix code. Then $Q^{(i)}$ is a prefix code by Lemma 3.1. But $Q^{(i)}$ is not a prefix code. This is a contradiction. So $AQ^{(i)}$ is not a prefix code.

Take $x \in A$ and $f \in Q$. Then $f^ix \in Q^{(i)}A$. Let $u = f^ix \in X^+$. By Lemma 3.2, for two distinct letters $a,\ b \in X$, at least one of the words ua and ub is primitive. Without loss of generality, we let ua is a primitive word that is, $f^ixa = ua \in Q$. So $(f^ixa)^ix \in Q^{(i)}A$. Since $f^ix,\ (f^ixa)^ix \in Q^{(i)}A$ and $f^ix \leq_p (f^ixa)^ix$, then $Q^{(i)}A$ is not a prefix code.

Assume $AQ^{(i)}B = A(Q^{(i)}B)$ is a prefix code. Then $Q^{(i)}B$ is a prefix code by Lemma 3.1. But $Q^{(i)}B$ is not a prefix code. This is a contradiction. So $AQ^{(i)}B$ is not a prefix code. □

We usually take the language $\{1\}$ as a prefix code. The family of all prefix codes over X is denoted by $P(X)$.

Lemma 3.4. *Let $u \in X^+$ and $lg(u) = k$. Then $ab^su,\ ba^su \in Q$ for all $s \geq k$ and $a,\ b \in X$ such that $a \neq b$.*

Proof. Assume ab^su is an imprimitive word. Then there exist $q \in Q$ and $t \geq 2$ such that $ab^su = q^t$. So $t \cdot lg(q) = s + 1 + k$. Then $2lg(q) \leq t \cdot lg(q) = s + 1 + k \leq s + 1 + s = 2s + 1$. So $lg(q) \leq \frac{2s+1}{2} < s + 1$. Hence $q <_p ab^s$. Then $q = ab^{s_1}$ for some $0 \leq s_1 < s$. So $ab^su = q^t = ab^{s_1}ab^{s_1} \ldots$. Hence $b^{s-s_1}u = abs_1 \ldots$. Thus $a = b$, which is a contradiction. Similarly, we can prove that ba^su is a primitive word. □

Lemma 3.5. (*Proposition 1.11 of* [5]) *Let $uv = f^i$ where $f \in Q$ and $i \geq 1$. Then $vu = g^i$ for some $g \in Q$.*

In this lemma, when $i = 1$, we can see if $uv \in Q$, then $vu \in Q$.

Proposition 3.6. *Let $A \in P(X) \setminus \{1\}$ and $B \in M$. Then $AQ^{(i)}B \neq Q^{(j)}$ for every $i,\ j \geq 2$.*

Proof. Assume there exist some $i,\ j \geq 2$ such that $AQ^{(i)}B = Q^{(j)}$. Take $x \in A$. Then $lg(x) = k \geq 1$. Let $a,\ b \in X$ and $a \neq b$. Then $ab^kx \in Q$ by Lemma 3.4. So $xab^k \in Q$ by Lemma 3.5. Then $(xab^k)^j \in Q^{(j)}$. Since $Q^{(j)} = AQ^{(i)}B$, then $(xab^k)^j \in AQ^{(i)}B$. There exist $f \in A$, $y \in Q$ and $z \in B$ such that $(xab^k)^j = fy^iz$. Since $A \in P(X)$ and $x,\ f \in A$, then $x = f$. Thus $(ab^kx)^{j-1}ab^k = y^iz$. In addition, exactly one of the following two cases may occur.

(1) When $lg(ab^kx) = lg(y)$, we have $ab^kx = y$. Since $(ab^kx)^{j-1}ab^k = y^iz$, then $(ab^kx)^{j-1}ab^k = (ab^kx)^iz$. Hence $(ab^kx)^{j-1}ab^k = (ab^kx)^{i-1}ab^kxz$. We consider the following case.

(1-1) When $i = j$, since $(ab^kx)^{j-1}ab^k = (ab^kx)^{i-1}ab^kxz$, then $(ab^kx)^{i-1}ab^k = (ab^kx)^{i-1}a$ b^kxz. Hence $xz = 1$. So $x = 1$. This contradicts that $lg(x) = k \geq 1$.

(1-2) When $i > j$, since $(ab^kx)^{j-1}ab^k = (ab^kx)^{i-1}ab^kxz$, then $ab^k = ab^kx(ab^kx)^{i-j}z$. So $x(ab^kx)^{i-j}z = 1$. Hence $x = 1$. This contradicts that $lg(x) = k \geq 1$.

(1-3) When $i < j$, since $(ab^kx)^{j-1}ab^k = (ab^kx)^{i-1}ab^kxz$, then $(ab^kx)^{j-i-1}ab^k = z$. Since $x \in A$, $ba^kx \in Q$ and $(ab^kx)^{j-i-1}ab^k = z \in B$, then $x(ba^kx)^i(ab^kx)^{j-i-1}ab^k = x(ba^kx)^iz \in AQ^{(i)}B$. Again since $AQ^{(i)}B = Q^{(j)}$, then $x(ba^kx)^i(ab^kx)^{j-i-1}ab^k \in Q^{(j)}$. There exists $g \in Q$ such that $x(ba^kx)^i(ab^kx)^{j-i-1}ab^k = g^j$. So $(xba^k)^i(xab^k)^{j-i}$ $= g^j$. In fact, we know $lg(xba^k) = lg(xab^k) = 2k + 1$. Calculating the lengths of the two words in the equation, we have $lg(g) = 2k + 1$. So $g = xba^k = xab^k$. Hence $a = b$. This contradicts that $a \neq b$.

(2) When $lg(ab^kx) \neq lg(y)$, we have $y = (ab^kx)^tu$ for some $0 \leq t \leq j - 1$ and $u \in X^*$. When $u = 1$, then $y = (ab^kx)^t$. Since $y \in Q$, then $t = 1$. Hence $y = ab^kx$. We have a contradiction by the former case. Hence $u \in X^+$. Since $(ab^kx)^{j-1}ab^k = y^iz = yy^{i-1}z = (ab^kx)^tuy^{i-1}z$, then $(ab^kx)^{j-1-t}ab^k = uy^{i-1}z$. When $0 \leq t \leq j - 2$, then $u <_p ab^kx$. When $t = j - 1$, then $u <_p ab^k$. So $u <_p ab^kx$ or $u <_p ab^k$. Since $ab^k <_p ab^kx$, then we consider $u <_p ab^kx$. One of the following cases may occur.

(2-1) If $u = ab^m$ for some $0 \leq m < k$, then $y = (ab^kx)^tu = (ab^kx)^tab^m$. So $(ab^kx)^{j-1}ab^k$ $= y^iz = yy^{i-1}z = (ab^kx)^tab^my^{i-1}z$. Then $(ab^kx)^{j-1-t}ab^k = ab^my^{i-1}z$. It follows that $b^{k-m}x(ab^kx)^{j-2-t}ab^k = y^{i-1}z$. We obtain y starts with the letter b. But by $(ab^kx)^{j-1}ab^k = y^iz$, we see that y starts with the letter a. This is a contradiction.

(2-2) If $u = ab^k$, then $y = (ab^kx)^tu = (ab^kx)^tab^k$. So $(ab^kx)^{j-1}ab^k = y^iz = yy^{i-1}z = (ab^kx)^tab^ky^{i-1}z$. It follows that $(ab^kx)^{j-1-t}ab^k = ab^ky^{i-1}z$. So $x(ab^kx)^{j-2-t}ab^k = y^{i-1}z$. We obtain the $(k + 1)$th letter in y is a. But by $y = (ab^kx)^tu$, we see that the $(k + 1)$th letter in y is b. This is a contradiction.

(2-3) If $u = ab^kx_1$ for some $x_1,\ x_2 \in X^+$ such that $x = x_1x_2$, then $y = (ab^kx)^tu = (ab^kx)^tab^kx_1$. So $(ab^kx)^{j-1}ab^k = y^iz = yy^{i-1}z = (ab^kx)^tab^kx_1y^{i-1}z$. It follows that $(ab^kx)^{j-1-t}ab^k = ab^kx_1y^{i-1}z$. So $x_2(ab^kx)^{j-2-t}ab^k = y^{i-1}z$. We obtain the $(lg(x_2) + 1)$th letter in y is a. But by $y = (ab^kx)^tu$, we see that the $(lg(x_2) + 1)$th letter in y is b. This is a contradiction.

From all above, we know $AQ^{(i)}B \neq Q^{(j)}$ for every $i,\ j \geq 2$. □

We cite some results from some references which will be used in the proof of our next proposition.

Lemma 3.7. (*Lemma 3.9 of* [8]) *Let A be a maximal prefix code and $B,\ D \in M\backslash\{1\}$. Then $AQ^{(i)}B \neq Q^{(i)}D$ for any $i \geq 2$.*

Lemma 3.8. (*Proposition 2.13 of* [5]) *Let $A,\ B,\ C,\ D,\ E$ be languages and $AB = CD = E$. Then $\underline{A}\ \underline{B} = \underline{C}\ \underline{D} = \underline{E}$ and $lg(A) + lg(B) = lg(C) + lg(D) = lg(E)$.*

Proposition 3.9. *Let $A \in P(X)\setminus\{1\}$, and $B,\ D \in M\setminus\{1\}$. Then $AQ^{(i)}B \neq Q^{(j)}D$ for every $i, j \geq 2$.*

Proof. Assume there exist $i \geq 2$ and $j \geq 2$ such that $AQ^{(i)}B = Q^{(j)}D$. We consider the cases that $i = j$ and $i \neq j$.

(1) If $i = j$, we consider the cases that A is a maximal prefix code and A is not a maximal prefix code.

(1-1) If A is a maximal prefix code, by Lemma 3.7, we have a contradiction.

(1-2) If A is not a maximal prefix code, then there exists $u_1 \in X^+\setminus A$ such that $A\cup\{u_1\} \in P(X)$. For the word u_1, by Lemma 3.2, we have $u_1a \in Q$ or $u_1b \in Q$ for two distinct letters $a,\ b \in X$. Without loss of generality, we may assume $u_1a \in Q$. For any $d_1 \in D$, by the assumption $AQ^{(i)}B = Q^{(j)}D$, we have $(u_1a)^id_1 \in Q^{(i)}D = AQ^{(i)}B$. There exist $y \in A$, $p \in Q$ and $v_1 \in B$ such that $(u_1a)^id_1 = yp^iv_1$. Since $A \cup \{u_1\}$ is a prefix code and $u_1,\ y \in A \cup \{u_1\}$, then $u_1 = y$. Since $y \in A$, then $u_1 \in A$. This contradicts that $u_1 \in X^+\setminus A$.

Thus for every $i, j \geq 2$, if $i = j$, then $AQ^{(i)}B \neq Q^{(j)}D$.

(2) If $i \neq j$, by the assumption $AQ^{(i)}B = Q^{(j)}D$, we have $\underline{AQ^{(i)}}\underline{B} = \underline{Q^{(j)}}\underline{D}$ by Lemma 3.8. Take $u_2 \in A$, $d_2 \in \underline{D}$ and $k = 2max\{lg(u_2),\ lg(d_2)\}$. Since $k \geq 2lg(u_2)$, then $k \geq lg(u_2) + 1 = lg(au_2)$ for $a \in X$. By Lemma 3.4, we have $ab^k(au_2) \in Q$ for $b \in X$ and $a \neq b$. Hence $u_2ab^ka \in Q$ by Lemma 3.5. Then $(u_2ab^ka)^jd_2 \in Q^{(j)}D = AQ^{(i)}B$. There exist $x \in A$, $q \in Q$ and $v_2 \in B$ such that $(u_2ab^ka)^jd_2 = xq^iv_2$. Since A is a prefix code and $u_2,\ x \in A$, then $x = u_2$. Hence $(ab^kau_2)^{j-1}ab^kad_2 = q^iv_2$. In the following, we consider $lg(q)$ and $lg(ab^kau_2)$.

(2-1) If $lg(q) = lg(ab^kau_2)$, then $q = ab^kau_2$. We consider the following cases.

(2-1-1) When $i > j$, since $q^{j-1}ab^kad_2 = q^iv_2$, then $ab^kad_2 = q^{i-j+1}v_2$. So $d_2 = u_2q^{i-j}v_2$. Calculating the lengths of the two words in the equation, we have $k > lg(d_2) = lg(u_2)+(i-j)(2+k+lg(u_2))+lg(v_2) > k$. This is a contradiction.

(2-1-2) When $i < j$, since $q^{j-1}ab^kad_2 = q^iv_2$, then $q^{j-i-1}ab^kad_2 = v_2 \in B$. By $ab^{k+2}u_2 \in Q$, we have $u_2(ab^{k+2}u_2)^iq^{j-i-1}ab^kad_2 \in AQ^{(i)}B = Q^{(j)}D$. There exist $g \in Q$ and $z \in D$ such that $u_2(ab^{k+2}u_2)^iq^{j-i-1}ab^kad_2 = g^jz$. Since $d_2 \in \underline{D}$, then $lg(d_2) \leq lg(z)$. So $z = z_1d_2$ for some $z_1 \in X^*$. It follows that $u_2(ab^{k+2}u_2)^iq^{j-i-1}ab^ka = g^jz_1$. So $lg(g) = 2 + k + lg(u_2) + \frac{i-lg(z_1)}{j}$. If $i - lg(z_1) \neq 0$, then $0 < \frac{i-lg(z_1)}{j} < 1$. This contradicts that $lg(g)$ is an integer. Hence $i - lg(z_1) = 0$. Then $q = u_2ab^{k+1} = bu_2ab^k$. We can see that the $(lg(u_2)+2)$th letter in q is a. On the other hand, we know $2(k+1)-(k+4) = 2k+2-k-4 = k-2 \geq 0$ because $k = 2max\{lg(u_2),\ lg(d_2)\} \geq 2$. So $k+1 \geq \frac{k+4}{2} = \frac{k}{2}+2$. Since $ab^kau_2 = q$ and $3 \leq lg(u_2)+2 \leq (\frac{k}{2})+2 \leq k+1$, then the $(lg(u_2)+2)$th letter in q is b. This is a contradiction.

(2-2) If $lg(q) \neq lg(ab^kau_2)$, let $q = (ab^kau_2)^tx_1$ for some $0 \leq t \leq j-1$ and $x_1 \in X^*$. Assume $x_1 = 1$, then $q = (ab^kau_2)^t$. Since $q \in Q$, then $t = 1$. We have a contradiction by case (2-1). Hence $x_1 \in X^+$. We consider the following cases.

(2-2-1) When $0 \leq t < j-1$, we have $x_1 <_p ab^kau_2$. One of the following cases may occur.

(2-2-1-1) If $x_1 = ab^{s_1}$ for some $0 \leq s_1 < k$, then $(ab^kau_2)^{j-1}ab^kad_2 = q^iv_2 = qq^{i-1}v_2 = (ab^kau_2)^tab^{s_1}q^{i-1}v_2$. So $(ab^kau_2)^{j-1-t}ab^kad_2 = ab^{s_1}q^{i-1}v_2$. Then $b^{k-s_1}au_2(ab^kau_2)^{j-2-t}ab^kad_2 = q^{i-1}v_2$. Hence q starts with the letter b. But by $q = (ab^kau_2)^tx_1$, we know that q starts with the letter a. This is a contradiction.

(2-2-1-2) If $x_1 = ab^k$, then we have $(ab^kau_2)^{j-1}ab^kad_2 = q^iv_2 = qq^{i-1}v_2 = (ab^kau_2)^tab^kq^{i-1}v_2$. So $(ab^kau_2)^{j-1-t}ab^kad_2 = ab^kq^{i-1}v_2$. It follows that $au_2(ab^kau_2)^{j-2-t}ab^kad_2 = q^{i-1}v_2$. Hence the $(lg(u_2)+2)$th letter in q is a. But by $q = (ab^kau_2)^tx_1$, we know that the $(lg(u_2)+2)$th letter in q is b because $(lg(u_2)+2) \leq k+1$. This is a contradiction.

(2-2-1-3) If $x_1 = ab^ka$, then we have $(ab^kau_2)^{j-1}ab^kad_2 = q^iv_2 = qq^{i-1}v_2 = (ab^kau_2)^tab^kaq^{i-1}v_2$. So $(ab^kau_2)^{j-1-t}ab^kad_2 = ab^kaq^{i-1}v_2$. It follows that $u_2(ab^kau_2)^{j-2-t}ab^kad_2 = q^{i-1}v_2$. Hence the $(lg(u_2)+1)$th letter in q is a. On the other hand, we know $2(k+1)-(k+4) = 2k+2-k-4 = k-2 \geq 0$ because $k = 2max\{lg(u_2),\ lg(d_2)\} \geq 2$. So $k+1 \geq \frac{k+4}{2} = \frac{k}{2}+2$. By $q = (ab^kau_2)^tx_1 = (ab^kau_2)^tab^ka$ and $3 \leq lg(u_2)+2 \leq (\frac{k}{2})+2 \leq k+1$, we can see that $(lg(u_2)+2)$th letter in q is b. This is a contradiction.

(2-2-1-4) If $x_1 = ab^kau_2'$ for some u_2', $u_2'' \in X^+$ such that $u_2 = u_2'u_2''$, then we have $(ab^kau_2)^{j-1}ab^kad_2 = q^iv_2 = qq^{i-1}v_2 = (ab^kau_2)^tab^kau_2'q^{i-1}v_2$. So $(ab^kau_2)^{j-1-t}ab^kad_2 = ab^kau_2'q^{i-1}v_2$. Then $u_2''(ab^kau_2)^{j-2-t}ab^kad_2 = q^{i-1}v_2$. Hence the $(lg(u_2'')+1)$th letter in q is a. But by $q = (ab^kau_2)^tx_1$, we know that the $(lg(u_2'')+1)$th letter in q is b because $(lg(u_2'')+1) \leq k+1$. This is a contradiction.

(2-2-2) When $t = j-1$, we have $q = (ab^kau_2)^{j-1}x_1$ and $x_1 <_p ab^kad_2$. This implies that $(ab^kau_2)^{j-1}ab^kad_2 = q^iv_2 = ((ab^kau_2)^{j-1}x_1)^iv_2$. So $ab^kad_2 = x_1((ab^kau_2)^{j-1}x_1)^{i-1}v_2$. Calculating the lengths of the two words in the equation, we obtain $k+2+lg(d_2) = i\cdot lg(x_1)+(j-1)(i-1)(k+2+lg(u_2))+lg(v_2)$. So $i\cdot lg(x_1)+(ij-i-j)(k+2+lg(u_2))+lg(v_2)-lg(d_2)-k-2 = 0$. This contradicts that $lg(d_2) \leq \frac{k}{2}$.

Thus for every $i, j \geq 2$, if $i \neq j$, then $AQ^{(i)}B \neq Q^{(j)}D$.

From all above, we know $AQ^{(i)}B \neq Q^{(j)}D$ for any $i, j \geq 2$. □

Lemma 3.10. (*Lemma 3.6 of* [8]) *Let B, $D \in M$. If there exist i, $j \geq 2$ such that $Q^{(i)}B = Q^{(j)}D$, then $Q^{(i)} = Q^{(j)}$ and $B = D$.*

Since $P(X)$ is free, then $P(\overset{\circ}{X})$ is a code and it is the unique irreducible generating set of $P(X)$. So a prefix code A in $P(\overset{\circ}{X})$ cannot be represented in the form $A = BC$ with $B,\ C \in P(X)$. We have the following theorem on $P(\overset{\circ}{X}) \cup \{Q^{(i)} \mid i = 2,\ 3,\ 4,\ \ldots\}$.

Theorem 3.11. *The family $P = P(\overset{\circ}{X}) \cup \{Q^{(i)} \mid i = 2,\ 3,\ 4,\ \ldots\}$ is a code, so P^* is a free monoid properly containing all prefix codes.*

Proof. Let $A_1A_2\cdots A_m = B_1B_2\cdots B_n$ for any $A_1,\ A_2,\ \ldots,\ A_m,\ B_1,\ B_2,\ \ldots,\ B_n \in P$. We will show that $m = n$ and $A_i = B_i$ for $i = 1, 2, \ldots, n$. We prove the theorem by induction on m.

(1) If $m = 1$, then $A_1 = B_1B_2\cdots B_n$.

(1-1) If A_1 and B_1 are both in $P(\overset{\circ}{X})$, then $A_1 = B_1$ and $B_2\cdots B_n = \{1\}$. So $m = n$ and $A_1 = B_1$.

(1-2) If A_1 and B_1 are both in $\{Q^{(i)} \mid i = 2,\ 3,\ 4,\ \ldots\}$, there exist $i_1,\ j_1 \geq 2$ such that $Q^{(i_1)} = Q^{(j_1)}B_2\cdots B_n$. By Lemma 3.10, we have $Q^{(i_1)} = Q^{(j_1)}$ and $B_2\cdots B_n = \{1\}$. So $m = n$ and $A_1 = B_1$.

(1-3) If $A_1 \in P(\overset{\circ}{X})$ and $B_1 \in \{Q^{(i)} \mid i = 2,\ 3,\ 4,\ \ldots\}$, then $A_1 = Q^{(j_2)}B_2\cdots B_n$ for some $j_2 \geq 2$. By Proposition 3.2, we know that $Q^{(i)}B$ is not a prefix code. But $A_1 \in P(\overset{\circ}{X}) \subseteq P(X)$. This is a contradiction.

(1-4) If $A_1 \in \{Q^{(i)} \mid i = 2,\ 3,\ 4,\ \ldots\}$ and $B_1 \in P(\overset{\circ}{X})$, then $Q^{(i_2)} = B_1B_2\cdots B_n$ for some $i_2 \geq 2$. If none of B_p belongs to $\{Q^{(i)} \mid i = 2,\ 3,\ 4,\ \ldots\}$ where $p \in \{2,\ 3,\ \ldots,\ n\}$, then $B_1B_2\cdots B_n$ is a prefix code. But $Q^{(i_2)}$ is not a prefix code. This is a contradiction. If some of B_p are in $\{Q^{(i)} \mid i = 2,\ 3,\ 4,\ \ldots\}$ where $p \in \{2,\ 3,\ ,\ldots,\ n\}$, let $B_t = Q^{(j_3)}$ for some $j_3 \geq 2$ and $2 \leq t \leq n$ such that B_t is the first one of $B_2,\ B_3,\ \ldots,\ B_n$ in $\{Q^{(i)} \mid i = 2,\ 3,\ 4,\ \ldots\}$. Then $Q^{(i_2)} = B_1B_2\cdots B_{t-1}Q^{(j_3)}B_{t+1}\cdots B_n$ where $B_1,\ B_2,\ \ldots,\ B_{t-1} \in P(\overset{\circ}{X})$. By Proposition 3.6, we know that this cannot be true.

(2) Assume that the theorem is true for integers proper less than m. For the integer m, we let $A_1A_2\cdots A_m = B_1B_2\cdots B_n$. We consider the following cases.

(2-1) If A_1 and B_1 are both in $P(\overset{\circ}{X})$, then $A_1 = B_1$ and $A_2\cdots A_m = B_2\cdots B_n$. According to the assumption, we have $m - 1 = n - 1$ and $A_i = B_i$ for $i = 2,\ 3,\ \ldots,\ m$. So $m = n$ and $A_i = B_i$ for $i = 1,\ 2,\ 3,\ \ldots,\ m$.

(2-2) If A_1 and B_1 are both in $\{Q^{(i)} \mid i = 2,\ 3,\ 4,\ \ldots\}$, then $Q^{(i_3)}A_2\cdots A_m = Q^{(j_4)}B_2\cdots B_n$ for some $i_3,\ j_4 \geq 2$. By Lemma 3.10, we have $Q^{(i_3)} = Q^{(j_4)}$, then $A_2\cdots A_m = B_2\cdots B_n$. According to the assumption, we know that the result holds.

(2-3) If $A_1 \in P(\overset{\circ}{X})$ and $B_1 \in \{Q^{(i)} \mid i = 2,\ 3,\ 4,\ \ldots\}$, then $A_1A_2\cdots A_m = Q^{(j_5)}B_2\cdots B_n$ for some $j_5 \geq 2$. If some of $A_2,\ A_3,\ \ldots,\ A_m$ are in $\{Q^{(i)} \mid i = 2,\ 3,\ 4,\ \ldots\}$, let $A_s = Q^{(i_4)}$ for some $i_4 \geq 2$ such that A_s is the first one of $A_2,\ A_3,\ \ldots,\ A_m$ in $\{Q^{(i)} \mid i = 2,\ 3,\ 4,\ \ldots\}$. Then $A_1A_2\cdots A_{s-1}Q^{(i_4)}A_{s+1}\cdots A_m = Q^{(j_5)}B_2\cdots B_n$ where $A_1,\ A_2,\ \ldots,\ A_{s-1} \in P(\overset{\circ}{X})$. Since $A_1A_2\cdots A_{s-1}$ is a prefix

code, by Proposition 3.9, this is impossible. If none of A_2, A_3, ..., A_m is in $\{Q^{(i)} \mid i = 2, 3, 4, \ldots\}$, then $A_1A_2\cdots A_m$ is a prefix code. By proposition 3.3, we know that $Q^{(j_5)}B_2\cdots B_n$ is not a prefix code. Since $A_1A_2\cdots A_m = Q^{(j_5)}B_2\cdots B_n$, this is a contradiction. Similarly, we can discuss the case $A_1 \in \{Q^{(i)} \mid i = 2, 3, 4, \ldots\}$ and $B_1 \in P(\mathring{X})$.

In summary, we know $P = P(\mathring{X}) \cup \{Q^{(i)} \mid i = 2, 3, 4, \ldots\}$ is a code. So P^* is a free monoid containing all the prefix codes in connection with the sets of ith powers of primitive words for all $i \geq 2$. □

Declarations

Conflict of interest The authors declare no conflict of interest.

References

[1] R Johannesson. *Informations theorie*, Addison-Wesley, 1992.

[2] J Karhumäki, L Michel, P Ion. *Commutation with codes*, Theoretical Computer Science, 2005, 340: 322-333.

[3] S Perkins, A E Escott. *Extended synchronizing code words for q-ary complete prefix codes*, Discrete Mathematics, 2001, 231: 391-401.

[4] M P Schützenberger. *On an application of semigroup methods to some problems in coding*, IRE Transactions on Information Theory, 1956, IT-2: 47-60.

[5] H J Shyr. *Free monoids and languages*, Hon Min Book Company, Taichung, Taiwan, China, 2001.

[6] H J Shyr. *Some free submonoids of the free monoid of prefix codes*, Semigroup Forum, 1975, 11: 22-29.

[7] J L Lassez, H J Shyr. *Factorization in the monoid of languages*, International Conference on Combinatorial Theory Canberra, International Mathematical Union, Australian Academic of Science, Lecture Notes in Mathematics, Springer-Verlag, 1977, 686: 229-235.

[8] H J Shyr, Y S Tsai. *Free submonoids in the monoid of languages*, Discrete Mathematics, 1998, 181: 213-222.

[9] J Bestel, D Perrin, C Reutenauer. *Codes and automata*, Cambridge University Press, 2009.

[10] S S Yu. *Languages and codes*, Taichung: Tsang Hai Book Publishing Company, 2005.

[11] G Păun, N Santean, G Thierrin, S Yu. *On the robustness of primitive words*, Discrete Applied Mathematics, 2002, 117: 239-252.

[1]School of Mathematics and Statistics, Yunnan University, Kunming 650405, China.

[2]Chongqing Tongnan Experimental Middle School, Chongqing 402660, China.

[3]Zhonghua Vocational College, Yunnan University of Finance and Economics, Kunming 650221, China.

Emails: yangdi65@aliyun.com, chhcao@ynu.edu.cn, 919762781@qq.com