ACTA MATHEMATICA VIETNAMICA
Volume 21, Number 2, 1996, pp. 279–289

A PROPERTY OF FINITE MAXIMAL CODES

NGUYEN HUONG LAM

Abstract. We establish a property of finite maximal codes relating exponent m such that b^m is in the code for some letter b to a structure pertaining to certain specific codewords. Consequently, we are able to propose some classes of finite codes that cannot be included in any finite maximal codes for almost all given exponents.

1. Introdution

In this article we give a characterization of finite maximal codes and use it to construct some classes of codes having no finite completions. For historical notes and background references we refer mostly to the book [1] and to the papers [2], [3], [4].

It is a simple, natural and well-expected fact that every code is included in a maximal code on the same alphabet which is called its completion: but it is nontrivial to discover that *not* every (finite) code is included in a *finite* maximal one. Some constructions of such codes are shown in [2], [3] and [4] all of which are based on a simple characterization of a special class of finite maximal codes. Maximal codes have a remarkable property that they are complete; and when a code is finite, completeness alone implies maximality. Thus it is easily seen that for each letter b of the alphabet a finite maximal code must contain b^m for some (unique) positive interger m. The construction starts by assuming beforehand for a code that it contains a given letter b (i.e. the exponent $m = 1$) and also a^n, for another letter a. Then it admits a so-called *unambiguous pair* of $\boldsymbol{Z}_n$ and when the code is finite maximal, the corresponding unambiguous pair turns out to be a factorization of $\boldsymbol{Z}_n$. The scope of construction is to search for finite codes with the associated unambiguous pair never completed to a factorization of $\boldsymbol{Z}_n$.

The aim of this paper is to extend this kind of characterization to any finite maximal code. Consequently, we can use it as a necessary condition to derive finitely incompletable codes for almost all pairs of exponents m and n. To do so, we utilize the familiar notion of multiset and come to a counterpart of the ordinary factorization: the multifactorization.

Received October 7, 1995

2. Multifactorization

First, we recall the concept of multiset. To avoid unnecessary formalism, we say simply that a *multiset* is any abstract set whose elements can enter several times. We often present a multiset M in an enumeration $M = \{e_1, e_2, \dots, e_{n-1}, e_n\}$ with e_i's not necessarily distinct. The number of occurences of a given element is the *multiplicity* of this element in the set. For example, the set $M = \{1, 2, 2, 3, 3, 3\}$ is a multiset of possitive intergers and the multiplicity of 3 is 3. Two multisets are considered equal if they contain the same elements, each with equal multiplicity. Further on, we shall use the traditional set-theoretic terminology whenever it is clear in the context that the multisets under consideration are those with all elements having multiplicity 1. Given a multisubset M, denote by $|M|$ the sum of all multiplicities of the elements in M; when M is the ordinary set, $|M|$ is simply the cardinality of M. For more information one may consult [6, Chapter VI].

Let $\boldsymbol{Z}_n$ be the additive group of residues modulo n: elements of $\boldsymbol{Z}_n$ are identified with their representatives in the complete system of residues $\{0, 1, \dots, n-1\}$ modulo n. Let H and K be multisets with elements taken from $\boldsymbol{Z}_n$, we say for convenience that H and K are *multisubsets* of $\boldsymbol{Z}_n$, in notation, $H \subseteq \boldsymbol{Z}_n$, $K \subseteq \boldsymbol{Z}_n$. The multisubset $H + K$ is to be defined as $\{h+k : h \in H, k \in K\}$ counting multiplicities. For instance, $H = \{0, 1, 1\}$, $K = \{2, 2, 2, 3, 3, 3, 3\} \subseteq \boldsymbol{Z}_4$, so $H + K = \{0+2, 0+2, 0+2, 0+3, 0+3, 0+3, 0+3, 1+2, 1+2, 1+2, 1+3, 1+3, 1+3, 1+3, 1+2, 1+2, 1+2, 1+3, 1+3, 1+3, 1+3\} = \{0, 0, 0, 0, 0, 0, 0, 0, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3\}$. We say also that a multiset is a *set of multiplicity* s if each element of it has multiplicity s.

Definition 2.1. A pair of multisubsets (H, K) of $\boldsymbol{Z}_n$ is said to be a *multifactorization of* $\boldsymbol{Z}_n$ *with multiplicity* s, or for short, s*-multifactorization* of $\boldsymbol{Z}_n$, if $H + K$ equals to the set $\boldsymbol{Z}_n$ of multiplicity s.

Speaking otherwise, it is the case when each residue modulo n is represented exactly s times as sum $h + k$ taken with multiplicity of h in H and k in K.

Example 2.1. The following multisubsets $H = \{0, 0, 1, 1, 4, 4, 5, 5\}$ and $K = \{0, 0, 0, 1, 2, 2, 2, 3\}$ of $\boldsymbol{Z}_8$ forms a 8-multifactorization of $\boldsymbol{Z}_8$.

Remark 2.1. 1-multifactorization ($s = 1$) is a factoroization in its traditional sense [2], [3].

3. Presentation of Results

Hereafter, we consider always a binary alphabet $A = \{a, b\}$ of two letters a and b; A^* is the free monoid of words on A with the concatenation as product. The symbol ε stands for the empty word (or the unit) of A^*. The notation $|w|$, for a word w, means the length of w. Given a subset X of A^*, we denote by X^* the Kleene closure of X and $X^+ = X^* - \{\varepsilon\}$. A subset C of A^* is a *code* if the equality

$$c_1 \dots c_i = d_1 \dots d_j$$

with $c_1, \dots, c_i,\ d_1 \dots, d_j \in C$ implies $i = j$ and $c_1 = d_1, \dots, c_i = d_i$. A code is called *maximal* if every strictly larger set is not a code. Remind also that a subset X is said to be *complete*, provided for all word w of A^*, $A^*wA^* \cap X^* \neq \emptyset$.

Given a finite maximal code C, there must be two possitive intergers n and m for which a^n and b^m belong to C. Let

$$P = C \cap b^*a^+ = \{b^{s_1}a^{t_1}, \dots, b^{s_p}a^{t_p}\}$$

and

$$Q = C \cap a^+b^* = \{a^{u_1}b^{v_1}, \dots, a^{u_q}b^{v_q}\},$$

where s_i, t_i, u_j, v_j are nonnegative intergers. Define the following multisubsets of $\boldsymbol{Z}_n$

$$R = \{t_1, \dots, t_p\} \bmod n$$

and

$$L = \{u_1, \dots, u_q\} \bmod n.$$

The main result is the following.

Theorem 3.1. *For any finite maximal code C over the binary alphabet $A = \{a, b\}$ with a^n, $b^m \in C$, the corresponding pair of multisubsets (R, L) constitutes an m-multifactorization of $\boldsymbol{Z}_n$.*

Remark 2. When $m = 1$, it is easily seen that R and L are the ordinary set and Theorem 3.1 asserts that (R, L) is a 1-multifactorization, i.e. a factorization of $\boldsymbol{Z}_n$. This is the characterization given in [3], [4].

The proof will be given in the next section with the decisive moment singled out as an independent proposition.

4. Proof of the main result

Let f be any word and X any subset of A^*. A word w is called *f-trim* (relative to the letter b and the set X, to be precise) if w can be presented in the form

$$w = b^i f b^j = x_1 \dots x_k$$

so that $x_1, \dots, x_k \in X$ and $0 \leq i = |b^i| < |x_1|$, $0 \leq j = |b^j| < |x_k|$.

It should be noted that an equality $b^j = b^k f b^\ell$ is possible for two pair (i,j) and (k,ℓ) distinct iff the occurences of f in the left and in the right sides overlap, that is iff $f \in b^*$. Therefore, if f contains an occurence of the letter a, i.e. $f \in A^*aA^*$, every f-trim word w is determined uniquely by the pair (i,j). Denote the set of f-trim words by $T(f)$.

The following proposition, originated from [5] and presented partially as Exercise 6.3, Chapter VIII, of [1], will be a central argument in this section.

Proposition 4.1. *Let C be a finite subset of A^* satisfying $C \cap b^* = \{b^m\}$ and f be an arbitrary word of A^*aA^*.*

(i) *If C is complete, the number of f-trim words is not less than m.*

(ii) *If C is a code, the number of f-trim words is not more than m.*

(iii) *If C is a maximal code, the number of f-trim words is exactly m.*

Proof of Proposition 4.1. (i) Assume for the contrary that the number of f-trim words is $< m$. Let r be an interger larger than twice the maximum of the lengths of the words of C and $b^x f b^r f b^y$ be an $fb^r f$-trim word. In the representation as a product of words of C

$$b^x f b^r f b^y = c_1 \dots c_k$$

with $x < |c_1|$, $y < |c_k|$, let i be the smallest interger such that $c_1 \dots c_i \geq |b^x f|$ and j be the largest one such that $|c_1 \dots c_j| \leq |b^x f b^r|$. Hence, by the choice of the large r, $c_1 \dots c_i = b^x f b^y$ and $c_{j+1} \dots c_k = b^q f b^y$ with $|c_i| > p > 0$, $|c_j| > q \geq 0$ meaning that $b^x f b^p$ and $b^q f b^y$ both are f-trim words. Also, $i < j$ and since b^m is the unique word in $C \cap b^*$, it follows $b_{i+1} = \cdots = b_j = b^m$ and therefore $r \equiv p + q \mod m$. Conversely, given any pair of f-trim words $b^x f b^p$ and $b^q f b^y$, for large $r \equiv p + q \mod m$ $(r = p + q + km)$, $b^x f b^r f b^y = b^x f b^p (b^m)^k b^q f b^y$ is indeed an $fb^r f$-trim word. Remind that as $f \notin b^*$, it is represented uniquely by the pair (x, y). This means that the total number of $fb^r f$-trim words when r runs through a fixed complete system of residues modulo m, does not exceed the number of pairs of f-trim words, thus is $\leq (m-1)^2$. Consequently, there exists a residue r such that the number of fb^rf-trim words is less or equal $\dfrac{(m-1)^2}{m} < m-1$. It is known that the number of f-trim words $< m$ implies the number of hb^rf-trim words $< m-1$ for some r. By descent, we comes to conclusion that there exists a word f', for which the number of f'-trim words is 0, but this obviously contradicts the completness of C.

(ii) Let C be a code and suppose on the contrary that the number of f-trim words equals $i > m$. Let further r be an integer sufficiently large and $b^xfb^rfb^y$ be any fb^rf-trim word. By the argument just used above, as C is code, we have a unique pair of f-trim words (b^xfb^p, b^qfb^y) satisfying $p+q \equiv \bmod m$. Conversly, any pair of f-trim words (b^xfb^p, b^qfb^y) yields an fb^rf-trim word $b^xfb^rfb^y$ with $r \equiv p+q \bmod m$. It follows that the total number of fb^rf-trim words when r runs through a fixed complete system of residues modulo m is exactly the number of distinct pairs of f-trim words, that is $i^2 > mi$. Hence the number of fb^rf-trim words for some residue r is $> i = \dfrac{mi}{m}$ and, indeed, $fb^rf \notin b^*$. Now by "ascent" argument, we conclude that there exists a word f' for which the number of f'-trim words is $> m^2$. But then we can single out two distinct f'-trim words $b^if'b^j$ and $b^kf'b^\ell$ so that $i \equiv k \bmod m$ and $j \equiv \ell \bmod m$. Because C is a code, this is a contradiction with the fact that $b^if'b^j$ and $b^kf'b^\ell$ are f'-trim words.

(iii) is straightforward from (i) and (ii): a maximal code is complete. The proof of Proposition 4.1 is complete.

Now we turn to the proof of the theorem 3.1. We fix a complete system of residues modulo n, each of them is sufficiently large by value and define $F_r = \{(d,e) \in (P,Q) : d \in b^*a^i,\ e \in a^jb^*,\ i+j \equiv r \bmod n\}$ for a residue r. We show that $|F_r| = m$. Put $f_r = a^r$. By Proposition 1 the f_r-trim words are m in number because $f_r \notin b^*$. Let $b^xf_rb^y$ be an arbitrary f_r-trim word, we have

$$b^xf_rb^y = c_1 \dots c_k$$

with $c_1, \dots, c_k \in C$; $x < |c_1|$, $y < |c_k|$. This implies $c_1 = b^xa^i \in P$, $c_k = a^jb^y \in Q$, $i,j > 0$ and $c_2 = \cdots = c_{k-1} = a^n$ and $i+j \equiv r \bmod n$. Since C is a code, for a given f_r-trim word, the pair (c_1, c_k) is unique, what we denote $c_1 = d(x,y)$, $c_k = e(x,y)$. Therefore the mapping $\theta_r : T(f_r) \to F_r$, given by

$$\theta_r(b^xf_rb^y) = (d(x,y), e(x,y))$$

for each f_r-trim word $w = b^x f_r b^y$, is well-defined. Moreover, it is injective.

On the other hand, this mapping is surjective. In fact, given any pair $(d, e) \in F_r$, $d = b^x a^i$, $e = a^j b^y$ for some $x, y \geq 0$, $i, j > 0$ such that $i + j \equiv r \mod n$, we have $r = kn + i + j$ as r is large enough, and $b^x a^i (a^n)^k a^j b^y = b^x a^r b^y = b^x f_r b^y$. Thus

$$\theta_r(b^x f_r b^y) = (d(x, y), e(x, y)) = (b^x a^i, a^j b^y) = (d, e),$$

showing the surjectivity of θ_r.

Summing up, θ_r is a bijection, and as consequence $|T(f_r)| = |F_r| = m$. Each $(d, e) \in F_r$ contributes one pair $(i, j) \in (R, L)$ such that $i + j \equiv r \mod n$ meaning that counting multiplicites, there are exactly m pair (i, j) of the multiset (R, L) with sum $i + j \equiv r \mod n$ what is to be proved.

Example 4.2. Code $C = \{a^2, b^2, b^2a, bab, a^2b, aba\}$ is maximal, $m = n = 2$; $L = \{2, 0\} = \{0, 0\}$ mode 2 and $R = \{1, 2\} = \{1, 0\}$ mod 2 and forms 2-multifactorization of $\boldsymbol{Z}_2$.

5. Examples: finitely incompletable codes

Theorem 3.1 provides a necessary condition for a finite maximal code. If a finite code containing a^n, b^m and admits the pair of multisets (P, Q) that is never completed to an m-multifactorization of $\boldsymbol{Z}_n$, the code itself cannot be completed to any finite maximal code either. Now we propose a construction of such codes for various values of m and n.

A positive integer n is said to have *property* P if there exist positive integers d, t and j such that $d > j$, $t > j$ and $n = dt + j$. A small arithmetical analysis will show that all positive integers have property P, except $n = 1, 2, 3, 4, 6, 8, 12, 16$ and 24.

Now given any integer m having property P, $m = dt + j$, $d > j > 0$, $t > j > 0$, and n any positive integer, let $H = \{h_1 = 0, \ldots, h_d\}$ and $K = \{k_1 = 0, \ldots, k_t\}$ form an unambigous pair (abbreviated u.a.p. henceforth) of $\boldsymbol{Z}_m$ [3]. This means that every residue of $\boldsymbol{Z}_m$ is representable in at most one way as a sum $h_i + h_j$ mod m. Let further $S = \{r_1 = 0, \ldots, r_p\}$ and $T = \{\ell_1 = 0, \ldots, \ell_q\}$ be a factorization of $\boldsymbol{Z}_n$, $(n = pq)$. Remind again that this means that every residues of $\boldsymbol{Z}_n$ is represented in exactly one way as a sum $\ell_i + r_j$ mod n. Consider the subset C_E consisting of the following words

$$\begin{array}{lr}
\{a^n & b^m \\
\;b^{h_2}a^n & a^n b^{k_2} \\
\;\ldots & \ldots \\
\;b^{h_d}a^n & a^n b^{k_t} \\
b^m a^{r_2} & a^{\ell_2} b^m \\
b^{h_2} a^{r_2} & a^{\ell_2} b^{k_2} \\
\ldots & \ldots \\
b^{h_d} a^{r_2} & a^{\ell_2} b^{k_t} \\
\ldots & \ldots \\
b^m a^{r_p} & a^{\ell_q} b^m \\
b^{h_2} a^{r_p} & a^{\ell_q} b^{k_2} \\
\ldots & \ldots \\
b^{h_d} a^{r_p} & a^{\ell_q} b^{k_t}\}.
\end{array}$$

One checks that C_E is a code, using, for example, the Sardinas-Patterson criterion and the fact that (H,K) and (S,T) are a.u.p. of $\boldsymbol{Z}_m$ and $\boldsymbol{Z}_n$, respectively. Clearly, the corresponding pair of multisubsets R, L of C_E are:

$$R = \{\underbrace{n,\ldots,n}_{d\,\text{times}};\underbrace{r_2,\ldots,r_2}_{d\,\text{times}};\underbrace{r_p,\ldots,r_p}_{d\,\text{times}}\} \bmod n$$

and

$$R = \{\underbrace{n,\ldots,n}_{d\,\text{times}};\underbrace{\ell_2,\ldots,\ell_2}_{d\,\text{times}};\underbrace{\ell_p,\ldots,\ell_p}_{d\,\text{times}}\} \bmod n$$

Evidently, (R,L) is not an m-multifactorization of $\boldsymbol{Z}_n$ and, moreover, we state that (R,L) is never completable to be so. As a matter of fact, for each element $x \in \boldsymbol{Z}_n$, there exists a pair $(r_i,\ell_j) \in (S,T)$ such that $x \equiv r_i + \ell_j \bmod n$. If we adjoin x to R, the pair $(R \cup \{x\}, L)$ represents the residue x in t ways by the sum $x+0$ and in dt ways by the sum $r_i+\ell_j$, as n is of multiplicity t, r_i of multiplicity d and ℓ_j of multiplicity t. Thus the multiplicity of x in $R \cup \{x\} + L$ is $dt + t > dt + j = m$. As far as adjoining x to L is concerned, by symmetry, the same remains true. So (R,L) cannot be made into an m-multifactorization of $\boldsymbol{Z}_n$ by adjoining any element to it. So C_E has no finite completions.

Example 5.3. For $m = 10$, $n = 2$ we choose $d = 3$, $t = 3$, $j = 1$, $p = 1$, $q = 2$. Take $H = \{0,1,2\}$, $K = \{0,3,6\}$ mod 10 and $S = \{0\}$, $T = \{0,1\}$ mod 2, the code $C_E = \{a^2, ba^2, b^2a^2, b^{10}, a^2b^3, a^2b^6, ab^{10}, ab^3, ab^6\}$.

This construction yields the desired result for all m, n unless both values are of $\{1, 2, 3, 4, 6, 8, 12, 16, 24\}$. We subsequently give a more refined example covering part of these exceptional values.

Let us say that an u.a.p. (H, K) of $\boldsymbol{Z}_n$ is maximal if for every u.a.p. (H', K') of $\boldsymbol{Z}_n$, $H \subseteq H'$ and $K \subseteq K'$ imply $H = H'$ and $K = K'$. For the sake of clarity, we reproduce here an auxiliary statement from [4].

Claim. *Given positive integers n, $d.t.$ j such that $t \geq 2$, $d > j$ and $n = dt + j$, let (H, K) be an u.a.p. of $\boldsymbol{Z}_n$ satisfying $K = \{0, 1, \ldots, d-1\}$, $\{0, d\} \subseteq H$ and $|H| = t$. Then (H, K) is a maximal u.a.p. of $\boldsymbol{Z}_n$.*

Proof of Claim. (Sketch.) For every $x \notin H \cup K$, the fact that $(H \cup \{x\}, K)$ is not an u.a.p. is evident. It will be shown that $(H, K \cup \{x\})$ is not an u.a.p. either. Let $h_0 = x_0 d + j_0, \ldots, h_{t-1} = x_{t-1} d + j_{t-1}$ be elements of H with $0 \leq j_0, \ldots, j_{t-1} < d$. Because $0, 1, \ldots, d-1 \in K$, it follows that

$$x_0 = 0,\ x_1 = 1, \ldots, x_{t-1} = t-1.$$

Put $x = rd + j'$, $j' < d$ $(r \leq t)$. Then the following three equalities show that $(H, K \cup \{x\})$ is not an u.a.p.

$$x + d = d + (j - j') \bmod n$$

if $r = t$, so that $j' < j$;

$$x = k_r + (j' - j_r)$$

if $r < t$ and $j' \geq j_r$; and

$$x + d = k_r + d - (j_r - j')$$

if $r < t$ and $j' < j_r$.

Explicit computation shows that each $n > 6$ satisfies the claim, that is there exists a maximal u.a.p. (H, K) of $\boldsymbol{Z}_n$ subject to the restrictions anounced in it. For such n and (H, K) and for every m with an integral factorization $m = pq$, let C_F to be defined as:

$$\begin{array}{ll} \{a^n & b^m \\ ba^n & a^n b^p \\ \cdots & \cdots \\ b^{p-1}a^n & a^n b^{p(q-1)}\} \end{array}$$

$$\bigcup_{h \in H, h \neq 0} \{b^m a^h, ba^h, \ldots, b^{p-1}a^h\} \quad \bigcup_{k \in K, k \neq 0} \{a^k b^m, a^k b^p, \ldots, a^k b^{p(q-1)}\}.$$

The codity of C_F is verified readily by Sardinas-Patterson criterion, taking the u.a.p. (H, K) into account. Evidently, the multisets R and L associated to C_E are the set H and K each element of which counted with the multiplicity p and q respectively: $R = \bigcup_{h \in H} \underbrace{\{h, \dots, h\}}_{p \text{ times}}$, $L = \bigcup_{k \in K} \underbrace{\{k, \dots, k\}}_{q \text{ times}}$. As (H, K) is not a factorization, (R, L) is not an m-multifactorization of $\boldsymbol{Z}_n$. If (R, L) were embedded into an m-multifactorization (R', L'), we may assume that there is some x in R' but not in R, thus not in H, as $H + K$ does not cover the whole $\boldsymbol{Z}_n$. The maximality of (H, K) implies that there exist $i \in \boldsymbol{Z}_n$, $h_1 \in H$, $k_1, k_2 \in K$. So that $i = h_1 + k_1 = x + k_2 \mod n$. But then the multiplicity of i in $R' + L'$ is at least pq (multiplicity of the sum $h_1 + k_1$) plus q (multiplicity of $x + k_2$) which is $\geq m + 1$: a contradiction. Thus (R, L) cannot be completed to any m-multifactorization of $\boldsymbol{Z}_n$ and consequently C_F has no finite completions.

This construction is possible for those m and n where at least one of them is not 1, 2, 3, 4 or 6.

Example 5.4. Let $m = 4$, $n = 8$, take $d = 3$, $t = 2$, $t = 2$; $p = q = 2$; $R = \{0, 3\}$, $L = \{0, 1, 2\}$. Then

$$C_F = \{a^8, b^4, ba^8, b^4a^3, ba^3, a^8b^2, ab^4, ab^2, a^2b^4, a^2b^2\}.$$

In closing, we should note that the few values m and n that remain are relatively small so that they could possibly lend themselves to an individual treatment.

Acknowledgment

I thank the referee for corrections and Do Long Van for useful disscussions.

References

1. J. Berstel, D. Perrin, *Theory of codes*, Academic Press, New York 1985.
2. C. De Felice, A. Restivo, *Some results on finite maximal codes*, RAIRO Informatique théorique **19** (1985), 383-403.
3. A. Restivo, S. Salemi, T. Sportelli, *Completing codes*, RAIRO Informatique théorique **23** (1989), 135-147.
4. N. H. Lam, *On codes having no finite completion*, Theoretical Informatics and Applications **29** (1995), 145-155.

5. D. Perrin, M. P. Schützenberger, *Codes et sous-monoïdes possédants des mots neutres*, Lecture Notes in Computer Science **48** (1977), 270-281.

6. S. Eilenberg, *Automata, Languages and Machines*, Volume **A**, Academic Press, New York, 1974.

Institute of Mathematics, P.O. Box 631, Bo Ho, Hanoi