2020 International Conference on Computer Science and Engineering Technology (CSET2020)

Bounds of Codes over Symbol-Pair Read Channels

Chengmin Wang

Department of Mathematics, Taizhou University, Taizhou, 225300, China

Email: chengmin_wang@163.com

Keywords: Symbol-Pair Code; Plotkin Type Bound; Johnson Type Bound; Cauchy-Schwartz Inequality

Abstract: -Symbol-pair read channels are first introduced by Cassuto and Blaum, in which the outputs of the read process are pairs of consecutive symbols. This new paradigm is motivated by the limitations of the reading process in some high density data storage systems. Soon later, a Singleton type bound for symbol-pair codes are established by Chee et al. and symbol-pair codes achieving this code are called MDS symbol-pair codes. With the benchmark, a lot of optimal symbol-pair are constructed by the effort of several papers. It is well known that bounds play an important role in coding theory. Two more bounds of symbol-codes are presented by Elishco et al very recently. In this paper, we continue the investigation of bounds of symbol-pair codes and establish two types of bounds, one is called Plotkin type bound and the other is call restricted Johnson type bound. We also present some examples of optimal symbol-pair codes that achieve these two new bounds.

1. Introduction

Symbol-pair coding theory has been widely studied due to its applications in channels where individual symbols cannot be read for some physical limitations. A coding-theoretic framework has been presented [1] to overcome pair-errors over symbol-pair read channels. Specially, it displayed a way making use of pair-vectors to characterize codewords. A relevant pair-distance metric has been set up and used to establish necessary and sufficient conditions for the pair-error model. Relationship between usual Hamming distance and pair-distance has been also been analyzed. Soon later, the Singleton type bound of symbol-pair code has been found [2] and its corresponding optimal codes are called MDS symbol-pair codes[3]. Since then, MDS symbol-pair codes have been studied widely and a lot of optimal such codes are constructed via different kind of techniques [6-9].

In addition, two other bounds are established in [4], one is called John type bound and the other is called linear programming bound. With these two bounds, some new optimal symbol-pair codes are presented. In coding theory, it is known that bounds play an important role because people can use them as criterions to measure the optimality of codes. Without bounds, we cannot judge the optimality of a code. In this paper, we will continue the investigation of codes for symbol-pair read channels and focus on two new types of bounds of such codes. One is Plotkin type bound and the other is restricted Johnson type bound[11].

2. Preliminaries

For a positive integer $n \geq 2$, $\mathbb{Z}_n$ denote the ring $\mathbb{Z}/n\mathbb{Z}$. Let Γ be a set of q elements, called *symbols*. Let Γ^n be the set of all n-length sequences over Γ. The coordinates of $X \in \Gamma^n$ are indexed by elements of $\mathbb{Z}_n$, so that $X = (x_0, x_1, \cdots, x_{n-1})$.

A *pair-vector* over Γ is a vector in $(\Gamma \times \Gamma)^n$. For any $X = (x_0, x_1, \cdots, x_{n-1}) \in \Gamma^n$, the *symbol-pair read vector* of X is the pair-vector (over Γ)

$$\pi(x) = ((x_0, x_1), (x_1, x_2) \cdots, (x_{n-2}, x_{n-1}), (x_{n-1}, x_0)).$$

Obviously, each vector $X \in \Gamma^n$ has a unique symbol-pair read vector $\pi(x) \in (\Gamma \times \Gamma)^n$.

For a vector $X \in \Gamma^n$, denote by $w_H(X)$ the Hamming weight of X. For any vector $X \in \Gamma^n$, the *pair-weight* of X is defined as

$$w_p(X) = w_H(\pi(x)).$$

A code $\mathbb{C}$ is said to ba a *constant pair-weight* if all codewords have the same pair-weight.

For two vectors $X, Y \in \Gamma^n$, denote by $d_H(X,Y)$ the Hamming distance between X and Y. The *pair-distance* between X and Y is defined as

$$d_p(X,Y) = d_H(\pi(X), \pi(Y))$$

A (q-ary) code of length n is a nonempty set $\mathbb{C} \subseteq \Gamma^n$. Define the minimum pair-distance of $\mathbb{C}$ as

$$d_p(\mathbb{C}) = \min\{d_p(X,Y) \mid X, Y \in \mathbb{C}, X \neq Y\}$$

A code $\mathbb{C}$ of length n over Γ is called an (n, M, d_p)-symbol-pair code if its size is M and the minimum pair-distance is d_p. Furthermore, if all the codewords of $\mathbb{C}$ have the same pair-weight w_p, we call $\mathbb{C}$ an (n, M, d_p, w_p) constant pair-weight symbol-pair code.

Let q, n, d_p, w_p be integers. Let $A_q(n, d_p)$ be the maximal size of a q-ary symbol-pair code of length n with pair-distance d_p and $A_q(n, d_p, w_p)$ be the maximal size of a q-ary symbol-pair code of length n with pair-distance d_p and constant weight w_p. We call a q-ary (n, d_p) symbol-pair code with size $A_q(n, d_p)$ is optimal. Similarly, a q-ary (n, d_p, w_p) constant pair-weight symbol-pair code with size $A_q(n, d_p, w_p)$ is called optimal.

3. The Plotkin Type Upper Bound

In this section, we present a Plotkin type upper bound. We notice that this type upper bound has been investigated in [10], in which the proof follows a similar logic as the proof of the Plotkin upper bound [5, Theorem 2.2.1]. Here we will provide the proof in detail for the convenience of readers. In addition, we will also improve the Plotkin bound strictly for the binary case.

Theorem 3.1. Let $\mathbb{C}$ be a q-ary (n, M, d_p) symbol-pair code over Γ such that $q^2 d_p > (q^2 - 1)n$. Then

$$A_q(n, d_p) \leq \left\lfloor \frac{q^2 d_p}{q^2 d_p - (q^2 - 1)n} \right\rfloor$$

Proof. Let

$$S = \sum_{X \in \mathbb{C}} \sum_{Y \in \mathbb{C}} d_p(X,Y)$$

If $X \neq Y$ for $X, Y \in \mathbb{C}$, then $d_p \leq d_p(X,Y)$ implying that

$$M(M-1)d_p \leq S \tag{1}$$

Let $\Re$ be the $M \times n$ matrix whose rows are the pair-vector of codewords of $\mathbb{C}$. For $1 \leq i \leq n$, let $n_{i,(\alpha,\beta)}$ be the number of times $(\alpha, \beta) \in \Gamma \times \Gamma$ occurs in columns i of $\Re$. As

$\sum_{(\alpha,\beta)\in\Gamma\times\Gamma} n_{i,(\alpha,\beta)} = M$ for $1 \le i \le n$, then we have

$$S = \sum_{i=1}^{n} \sum_{(\alpha,\beta)\in\Gamma\times\Gamma} n_{i,(\alpha,\beta)}(M - n_{i,(\alpha,\beta)}) = nM^2 - \sum_{i=1}^{n} \sum_{(\alpha,\beta)\in\Gamma\times\Gamma} n_{i,(\alpha,\beta)}^2 \quad (2)$$

By the Cauchy-Schwartz inequality, we have

$$\left(\sum_{(\alpha,\beta)\in\Gamma\times\Gamma} n_{i,(\alpha,\beta)} \right)^2 \le q^2 \sum_{(\alpha,\beta)\in\Gamma\times\Gamma} n_{i,(\alpha,\beta)}^2 \quad (3)$$

By (2) and (3), we obtain

$$S \le nM^2 - \sum_{i=1}^{n} q^{-2} \left(\sum_{(\alpha,\beta)\in\Gamma\times\Gamma} n_{i,(\alpha,\beta)} \right)^2 = n(1 - q^{-2})M^2 \quad (4)$$

Since M is an integer, combining (1) and (4), we obtain the following upper bound

$$M \le \left\lfloor \frac{q^2 d_p}{q^2 d_p - (q^2 - 1)n} \right\rfloor .$$

Then the proof is complete. □

If $q = 2$, the following result is clear.

Corollary **3.1.** Let $\mathbb{C}$ be a binary (n, M, d_p) symbol-pair code such that $4d_p > 3n$. Then

$$A_2(n, d_p) \le \left\lfloor \frac{4d_p}{4d_p - 3n} \right\rfloor$$

Example 1. Each row of the following array is a codeword of $\mathbb{C}$. It is readily verified that for the symbol-pair code $\mathbb{C}, n = 7, d_p = 6, M = 8$.

$$\mathbb{C} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Since

$$A_2(7,6) \le \left\lfloor \frac{4d_p}{4d_p - 3n} \right\rfloor = \left\lfloor \frac{4 \times 6}{4 \times 6 - 3 \times 7} \right\rfloor = 8,$$

$\mathbb{C}$ presented above is an optimal binary (7, 8, 6) symbol-pair code. □

Actually, the Plotkin type bound of symbol-pair codes for the binary case can be further improved.

Theorem 3.2. Let $\mathbb{C}$ be a binary (n, M, d_p) symbol-pair code over $\mathbb{Z}_2$ such that $4d_p > 3n$. Then

$$A_2(n,d_p) \le \begin{cases} 2\left\lfloor \dfrac{2d_p}{4d_p-3n} \right\rfloor & \text{if } A_2(n,d_p) \text{ is even,} \\ \left\lfloor \dfrac{3n}{4d_p-3n} \right\rfloor & \text{if } A_2(n,d_p) \text{ is odd.} \end{cases}$$

Proof. Let $\Re$ be the $M \times n$ matrix whose rows are the pair-vector of codewords of $\mathbb{C}$. By Corollary 3.1, we have $M \le \left\lfloor \dfrac{4d_p}{4d_p-3n} \right\rfloor$ for the binary case. If M is even, we can round the expression $\dfrac{4d_p}{4d_p-3n}$ down to the nearest even integer, which gives the result.

If M is odd, we do not use Cauchy-Schwartz in the proof of Theorem 3.1. Instead, from (2), we observe that

$$\begin{aligned} S &= \sum_{i=1}^{n}[n_{i,(0,0)}(M-n_{i,(0,0)})+n_{i,(0,1)}(M-n_{i,(0,1)})+n_{i,(1,0)}(M-n_{i,(1,0)})+n_{i,(1,1)}(M-n_{i,(1,1)})] \\ &= \sum_{i=1}^{n}[M(n_{i,(0,0)}+n_{i,(0,1)}+n_{i,(1,0)}+n_{i,(1,1)})-n_{i,(0,0)}^2-n_{i,(0,1)}^2-n_{i,(1,0)}^2-n_{i,(1,1)}^2] \\ &= \sum_{i=1}^{n}[M^2-n_{i,(0,0)}^2-n_{i,(0,1)}^2-n_{i,(1,0)}^2-n_{i,(1,1)}^2] \\ &= \sum_{i=1}^{n}[M^2-\sum_{(\alpha,\beta)\in\mathbb{Z}_2\times\mathbb{Z}_2} n_{i,(\alpha,\beta)}^2] \end{aligned}$$

Since, for each $1 \le i \le n$,

$$\begin{aligned} M^2 &= \left(\sum_{(\alpha,\beta)\in\mathbb{Z}_2\times\mathbb{Z}_2} n_{i,(\alpha,\beta)}\right)^2 \\ &= \sum_{(\alpha,\beta)\in\mathbb{Z}_2\times\mathbb{Z}_2} n_{i,(\alpha,\beta)}^2 + 2n_{i,(0,0)}n_{i,(0,1)} + 2n_{i,(0,0)}n_{i,(1,0)} + 2n_{i,(0,0)}n_{i,(1,1)} \\ &\quad + 2n_{i,(0,1)}n_{i,(1,0)} + 2n_{i,(0,1)}n_{i,(1,1)} + 2n_{i,(1,0)}n_{i,(1,1)} \end{aligned}$$

Then

$$\begin{aligned} S = \sum_{i=1}^{n}[&2n_{i,(0,0)}n_{i,(0,1)} + 2n_{i,(0,0)}n_{i,(1,0)} + 2n_{i,(0,0)}n_{i,(1,1)} \\ &+ 2n_{i,(0,1)}n_{i,(1,0)} + 2n_{i,(0,1)}n_{i,(1,1)} + 2n_{i,(1,0)}n_{i,(1,1)}] \end{aligned} \tag{5}$$

If $M \equiv 1$ (mod 4), the right-hand side of (5) is maximized when

$$\{n_{i,(0,0)}, n_{i,(0,1)}, n_{i,(1,0)}, n_{i,(1,1)}\} = \{\frac{M-1}{4}, \frac{M-1}{4}, \frac{M-1}{4}, \frac{M+3}{4}\}$$

Thus, combining (1) and (5), we obtain

$$M(M-1)d_p \le \frac{3n}{8}[(M-1)^2+(M-1)(M+3)] = \frac{3n}{4}[(M+1)(M-1)] \tag{6}$$

Similarly, if $M \equiv 3$ (mod 4), the right-hand side of (5) is maximized when

$$\{n_{i,(0,0)}, n_{i,(0,1)}, n_{i,(1,0)}, n_{i,(1,1)}\} = \{\frac{M+1}{4}, \frac{M+1}{4}, \frac{M+1}{4}, \frac{M-3}{4}\}$$

Thus we still obtain

$$M(M-1)d_p \le \frac{3n}{8}[(M+1)^2+(M+1)(M-3)]=\frac{3n}{4}[(M+1)(M-1)] \tag{7}$$

Hence, by (6) or (7), we obtain

$$M \le \frac{3n}{4d_p-3n}.$$

The proof is complete. □

Remark 1. By comparing Corollary 3.1 and Theorem 3.2, it is clear that when M is odd, the bound $M \le \frac{3n}{4d_p-3n}$ is strictly better than $M \le \frac{4d_p}{4d_p-3n}$ because $3n \le 4d_p$. It is a meaningful work to construct optimal binary symbol-pair codes which achieve the bound of Theorem 3.2.

4. The Restricted Johnson Type Bound

In this section, we consider another new upper bound, the restricted Johnson type bound, which is concern on the constant pair-weight symbol-pair codes. The proof follows the logic as the proof of the upper bound [5, Theorem 2.3.4].

Theorem 4.1. Let $\mathbb{C}$ be a q-ary (n,M,d_p,w_p) constant pair-weight symbol-pair code over Γ such that $q^2w_p^2-2(q^2-1)nw_p+nd_p(q^2-1)>0$, then

$$A_q(n,d_p,w_p) \le \left\lfloor \frac{nd(q^2-1)}{q^2w_p^2-2(q^2-1)nw_p+nd_p(q^2-1)} \right\rfloor$$

Proof. The second bound is a special case of the first one. The proof of the first one uses the same idea as in the proof of the Plotkin Bound. Let $\mathbb{C}$ be an (n,M,d_p,w_p) constant weight symbol-pair code. Let $\Re$ be the $M\times n$ matrix whose rows are the pair-vectors of the codewords of $\mathbb{C}$.

Let

$$S=\sum_{X\in\mathbb{C}}\sum_{Y\in\mathbb{C}} d_p(X,Y)$$

If $X \ne Y$ for $X,Y\in\mathbb{C}$, then $d_p \le d_p(X,Y)$ implying that

$$M(M-1)d \le S \tag{8}$$

For $1\le i\le n$, let $n_{i,(\alpha,\beta)}$ be the number of times $(\alpha,\beta)\in\Gamma\times\Gamma$ occurs in columns i of $\Re$. As $\sum_{(\alpha,\beta)\in\Gamma\times\Gamma} n_{i,(\alpha,\beta)}=M$ for $1\le i\le n$, we have

$$\begin{aligned} S &= \sum_{i=1}^{n}\sum_{(\alpha,\beta)\in\Gamma\times\Gamma} n_{i,(\alpha,\beta)}(M-n_{i,(\alpha,\beta)}) \\ &= \sum_{i=1}^{n}(Mn_{i,(0,0)}-n_{i,(0,0)}^2)+\sum_{i=1}^{n}\sum_{(\alpha,\beta)\in\Gamma\times\Gamma\setminus(0,0)}(Mn_{i,(\alpha,\beta)}-n_{i,(\alpha,\beta)}^2) \end{aligned} \tag{9}$$

We analyze each of the last two terms one by one.

First, because $\sum_{i=1}^{n} n_{i,(0,0)}$ counts the number of $(0,0)$'s in the matrix $\Re$ and each of the M rows of $\Re$ has $n-w_p$ $(0,0)$'s, we have

$$\sum_{i=1}^{n} n_{i,(0,0)}=(n-w_p)M$$

Second, by the Cauchy-Schwartz inequality,

$$\left(\sum_{i=1}^{n} n_{i,(0,0)}\right)^2 \le n\sum_{i=1}^{n} n_{i,(0,0)}^2$$

Combining these we see that the first summation on the right-hand of (9) satisfies

$$\begin{aligned}\sum_{i=1}^{n}(Mn_{i,(0,0)} - n_{i,(0,0)}^2) &\le (n-w_p)M^2 - \frac{1}{n}(\sum_{i=1}^{n} n_{i,(0,0)})^2 \\ &= (n-w_p)M^2 - \frac{(n-w)^2 M^2}{n}\end{aligned} \qquad (10)$$

For the second summation, we have

$$\sum_{i=1}^{n}\sum_{(\alpha,\beta)\in\Gamma\times\Gamma\backslash(0,0)} n_{i,(\alpha,\beta)} = w_p M$$

By the Cauchy-Schwartz inequality,

$$\left(\sum_{i=1}^{n}\sum_{(\alpha,\beta)\in\Gamma\times\Gamma\backslash(0,0)} n_{i,(\alpha,\beta)}\right)^2 \le n(q^2-1)\sum_{i=1}^{n}\sum_{(\alpha,\beta)\in\Gamma\times\Gamma\backslash(0,0)} n_{i,(\alpha,\beta)}^2$$

This produces

$$\begin{aligned}\sum_{i=1}^{n}(Mn_{i,(\alpha,\beta)} - n_{i,(\alpha,\beta)}^2) &\le w_p M^2 - \frac{1}{n(q^2-1)}\left(\sum_{i=1}^{n}\sum_{(\alpha,\beta)\in\Gamma\times\Gamma\backslash(0,0)} n_{i,(\alpha,\beta)}^2\right)^2 \\ &= w_p M^2 - \frac{1}{n(q^2-1)}(w_p M)^2\end{aligned} \qquad (11)$$

Combining (8),(9),(10) and (11), we obtain

$$M(M-1)d_p \le (n-w_p)M^2 - \frac{(n-w_p)^2 M^2}{n} + w_p M^2 - \frac{1}{n(q^2-1)}(w_p M)^2$$

By simplying, we get

$$A_q(n,d_p,w_p) \le \left\lfloor \frac{nd(q^2-1)}{q^2 w_p^2 - 2(q^2-1)nw_p + nd_p(q^2-1)} \right\rfloor$$

So the proof is complete.□

Example 2. Each row of the following array is a codeword of $\mathbb{C}$. It is readily verified that for the constant pair-weight symbol-pair $\mathbb{C}$, $n=7, d_p=6, w_p=5, M=7$.

$$\mathbb{C} = \begin{pmatrix} 1&0&1&1&0&0&0 \\ 0&1&0&1&1&0&0 \\ 0&0&1&0&1&1&0 \\ 0&0&0&1&0&1&1 \\ 1&0&0&0&1&0&1 \\ 1&1&0&0&0&1&0 \\ 0&1&1&0&0&0&1 \end{pmatrix}$$

Since

$$A_2(7,6,5) \le \left\lfloor \frac{3nd_p}{4w_p^2 - 6nw_p + 3nd_p} \right\rfloor = \left\lfloor \frac{3\times7\times6}{4\times5^2 - 6\times7\times5 + 3\times7\times6} \right\rfloor = \left\lfloor \frac{63}{8} \right\rfloor = 7$$

$\mathbb{C}$ presented above is an optimal binary (7, 8, 6, 5) constant pair-weight symbol-pair code.

Conclusion

In this paper, we study the bounds for symbol-pair codes. Previously, the Singleton Type bound has been found [2] and its corresponding optimal codes are called MDS symbol-pair codes which have been studied widely. A lot of optimal MDS symbol-pair codes are constructed via different kinds of techniques. In [4], two other bounds are established and some optimal symbol-codes are also presented. It is clear that to produce optimal symbol-codes with some parameters, we have to establish bounds first. In this paper, we establish two upper bounds for symbol-pair codes and some examples for optimal symbol-pair codes which achieve the upper bounds. It is meaningful to construct more classes optimal symbol-pair codes which achieve the bounds obtained in this paper.

Acknowledgment

The research is supported by Natural Science Foundation of Jiangsu (Grant No. BK20171318), Qing Lan Project of Jiangsu and Overseas Study Program of Jiangsu University.

Reference

[1]. Cassuto Y and Blaum M 2011, Codes for Symbol-Pair Read Channels, Trans. Inf. Theory 8011-8020

[2]. Chee Y M, Ji L, Kiah H M, Wang C and Yin J 2013, Maximum Distance Separable Symbol-Pair Codes, IEEE Trans. Inf. Theory 2886-2890

[3]. Ding B, Ge G, Zhang J, Zhang T and Zhang Y 2017, New Constructions of MDS Symbol-Pair Codes, Des. Codes Cryp. 1-177

[4]. Elishco O, Gabrysy R and Yaakobi E 2018, Bounds and Constructions of Codes over Symbol-Pair Read Channels, IEEE International Symposium on Information Theory 2505-2509

[5]. Huffman W and Pless V 2003, Fundermentals of Error Correcting Codes, Cambridge University Press

[6]. Li S and Ge G 2017, Constructions of Maximum Distance Separable Symbol-Pair Codes Using Cyclic and Constacyclic Codes, Des. Codes Cryp. 359-371

[7]. Kai X S, Zhu S X and Li P 2015, A Construction of New MDS Symbol-Pair Codes, IEEE Trans. Inf. Theory 5828-5834

[8]. Sun Z, Zhu S and Wang L 2017, The Symbol-Pair Distance Distribution of a Class of Repeated-root Cyclic Codes over F_{p^m}, Cryp. and Comm. 643-653

[9]. Yaakobi E, Bruck J and Siegel P H 2016, Construction and Decoding of Cyclic Codes over b -Symbol Read Channels, IEEE Trans. Inf. Theory 1541-1551

[10]. Yang M, Li J and Feng K 2016, Construction of Cyclic and Constacyclic Codes for b -symbol Read Channels Meeting the Plotkin-like Bound, arXiv:1607.02677

[11]. Zhang H 2017, Improvement on Minimum Distance of Symbol-Pair Codes, Cryptography and Coding 116-124