A New Bound for Difference Systems of Sets

Hao Wang*
Department of Mathematical Sciences,
Michigan Technological University,
Houghton, MI 49931-1295, USA
hwang@mtu.edu

Abstract

Difference systems of sets DSS, introduced by Levenshtein, are used to design code synchronization in the presence of errors. The paper gives a new lower bound of DSS's size.

1 Introduction

Comma-free codes were introduced by Crick, Griffith, and Orgel in a biology paper [2]. One year later, the paper "Comma-free codes" [4] by Golomb, Gordon, and Welch gave some essential mathematical results for these codes. To construct comma-free codes as cosets of linear codes, Levenshtein gave the definition of DSS in [6] as follows. A *difference system of sets* (DSS) with parameters $(n, \tau_0, ..., \tau_q - 1, \rho)$ is a collection of q disjoint subsets $Q_i \subseteq \{1, 2, ..., n\}, |Q_i| = \tau_i, 0 \leq i \leq q-1$, such that the multi-set

$$M = \{a - b \pmod{n} \mid a \in Q_i, b \in Q_j, i \neq j\} \tag{1}$$

contains every number i, $1 \leq i \leq n-1$, at least ρ times. A DSS is *perfect* if every number $i, 1 \leq i \leq n-1$, is contained exactly ρ times in the multi-set of differences (1). A DSS is *regular* if all subsets Q_i are of the same size. We use the notation (n, m, q, ρ) for a regular DSS on n points with q subsets of size m. Actually, Clague first studied DSS for the case $q = 2$ in [1] where he used the word "synchronous" rather than "comma-free".

Let F_q^n be the set of vectors of length n over $F_q = \{0, 1, ..., q-1\}$ for some positive integer q. Suppose $x = x_1...x_n$ and $y = y_1...y_n$ are two vectors. The *ith overlap* of x and y is defined as

$$T_i(x, y) = x_{i+1}...x_n y_1...y_i, \quad 1 \leq i \leq n-1.$$

*Research supported by NSF Grant CCR-0310832

Some scholars use *splice* or *joint* in lieu of the name "joint". Obviously, $T_i(x,x)$ is a cyclic shift of x. A code $C \subseteq F_q^n$ is called *comma-free* if any joint of two codewords is not a codeword in C. The *comma-free index* $\rho(C)$ of a code $C \subseteq F_q^n$ is defined as

$$\rho(C) = \min(z, T_i(x,y)),$$

where the minimum is taken over all $x, y, z \in C$ and all $i = 1, ..., n-1$, and d is the Hamming distance between vectors in F_q^n. The comma-free index $\rho(C)$ allows one to distinguish a code word from an overlap of two code words provided that at most $\lfloor \rho(C)/2 \rfloor$ errors have occurred in the given code word [4].

It is known that he comma-free index of any linear code is zero since any linear code contains the zero vector. Levenshtein [6] gave the following construction of comma-free codes of index $\rho > 0$ obtained as cosets of linear codes. Given a DSS $\{Q_0, ..., Q_{q-1}\}$ with parameters $(n, \tau_0, ..., \tau_{q-1}, \rho\}$, we define a linear code $C \subseteq F_q^n$ of dimension $n - r$, where

$$r = \sum_{i=0}^{q-1} |Q_i|.$$

For any vector $x \in C$, information positions are indexed by the numbers not contained in any of the sets $Q_0, ..., Q_{q-1}$ meanwhile we assign symbol i $(0 \leq i \leq q-1)$ to any position indexed by Q_i. This yields a coset C' of C that has a comma-free index at least ρ. From this construction method, it is desirable that the redundancy r is as small as possible.

Let $r_q(n, \rho)$ denote the minimum redundancy of a DSS with parameters n, q, and ρ. Levenshtein proved the following lower bound on $r_q(n, \rho)$ in [6].

Theorem 1.1

$$r_q(n, \rho) \geq \sqrt{\frac{q\rho(n-1)}{q-1}}, \tag{2}$$

with equality if and only if the DSS is perfect and regular.

In [6], Levenshtein also gave optimal DSS for $\rho = 1$ or 2 and $q = 2$, and proved that that for all $n \geq 2$

$$r_2(n,1) = \left\lceil \sqrt{2(n-1)} \right\rceil, \quad r_2(n,2) = \left\lceil 2\sqrt{n-1} \right\rceil.$$

A positive integer is called *square-free* if its prime decomposition contains no repeated factors. For example, 30 is square-free since its prime decomposition contains no repeated factors. The Möbius μ function $\mu(n)$ is defined as follows. For $n \in \mathbb{Z}^+$, $\mu(1) = 0$, $\mu(n) = 0$ if n is not square-free, and $\mu(p_1 p_2 \cdots p_l) = (-1)^l$, where the p_i are distinct positive primes. In this paper, we prove the following lower bound on $r_q(n, \rho)$.

Theorem *If* $\sqrt{\frac{q\rho(n-1)}{q-1}}$ *is a square-free integer, then*

$$r_q(n,\rho) \geq \sqrt{\frac{q\rho(n-1)}{q-1}} + 1.$$

Otherwise, $r_q(n,\rho) \geq \sqrt{\frac{q\rho(n-1)}{q-1}}$.

Recently, Levenshtein wrote a survey [7] on comma-free codes, which gives well-known results and methods, presents some new results and formulates open problems. Tonchev [8], [9] described some direct constructions of perfect and regular DSS as partitions of cyclic difference sets. Mutoh and Tonchev [5] gave several constructions of optimal DSS using cyclotomy. Cummings [3] gave another construction method for DSS and two easily expressed conditions for a systematic code to be comma-free.

2 A New Lower Bound on $r_q(n,\rho)$

Lemma 2.1 *If* $\{Q_0, Q_1, ..., Q_{q-1}\}$ *is a DSS with parameters* n *and* ρ*, then*

$$\sum_{i\neq j} \tau_i\tau_j \geq \rho(n-1). \tag{3}$$

Proof: Because $Q_0, ..., Q_{q-1}$ are disjoint subsets, the size of M is $\sum_{i\neq j}\tau_i\tau_j$ which has to be greater than or equal to $\rho(n-1)$ since each integer $1, ..., n-1$ must appear at least ρ times by definition. ∎

Lemma 2.2 *If a DSS is perfect then* $\rho(n-1)$ *is even.*

Proof: Since DSS is perfect, $\rho(n-1) = \sum_{i\neq j}\tau_i\tau_j = 2\sum_{i<j}\tau_i\tau_j$. ∎

A list of positive integers $[\tau_0, \tau_1, ..., \tau_{q-1}]$ is called a *q-partition* of r if $r = \sum_{i=0}^{q-1}\tau_i$. A partition is called a *fair* partition if $|\tau_i - \tau_j| \leq 1$ whenever $i \neq j$.

Lemma 2.3 *Suppose* $\{Q_0, Q_1, ..., Q_{q-1}\}$ *is a DSS with parameters* n, ρ*. Let* r *be the sum of sizes of* Q_i*, i.e.* $r = \sum_{i=0}^{q-1}|\tau_i|$*. If* $[\tau'_0, \tau'_1, ..., \tau'_{q-1}]$ *is a fair partition of* r*, then* $\sum_{i\neq j}\tau'_i\tau'_j \geq \rho(n-1)$.

Proof: Without loss of generality, suppose $\tau_0 \geq \tau_1 \geq ... \geq \tau_{q-1}$. Because $\{Q_i\}$ is a DSS, $\sum_{i\neq j}\tau_i\tau_j \geq \rho(n-1)$. If there are s and t $(s > t)$ such that $\tau_s \geq \tau_t + 2$,

then let $\tau_i' = \tau_i$ except for $\tau_s' = \tau_s - 1$ and $\tau_t' = \tau_t + 1$. Then because $\tau_s \geq \tau_t + 2$,

$$\begin{aligned} & \sum_{i \neq j} \tau_i' \tau_j' - \sum_{i \neq j} \tau_i \tau_j \\ = & \sum_{k \neq s,t} \tau_k(\tau_s - 1) + \sum_{k \neq s,t} \tau_k(\tau_t + 1) + (\tau_s - 1)(\tau_t + 1) \\ & - \sum_{k \neq s,t} \tau_k \tau_s - \sum_{k \neq s,t} \tau_k \tau_t - \tau_s \tau_t \\ = & \tau_s - \tau_t - 1 \geq 1. \end{aligned}$$

So we have

$$\sum_{i \neq j} \tau_i' \tau_j' > \sum_{i \neq j} \tau_i \tau_j \geq \rho(n-1).$$

Hence if there is a DSS with parameter r, then the inequality (3) must hold for any fair partition of r. ∎

Lemma 2.4 *The following inequality must be satisfied for any DSS:*

$$r(r-1) + (q-2r)\lfloor \frac{r}{q} \rfloor + q\lfloor \frac{r}{q} \rfloor^2 \geq \rho(n-1),$$

where $\lfloor x \rfloor$ is the greatest integer less than or equal to x. The equality holds if and only if the following two conditions are true.

(1) The DSS is perfect.

(2) The sizes of $\{Q_0, Q_1, ..., Q_{q-1}\}$ form a fair partition of r.

Proof: From lemma 2.3, we only need to calculate the size of multi-set M when $[\tau_0, \tau_1, ..., \tau_{q-1}]$ is a fair partition of r. $\lfloor x \rfloor$ and $\lceil x \rceil$ are used to denote the greatest integer less than or equal to x and the smallest integer greater than or equal to x, respectively. Let $r = q\lfloor \frac{r}{q} \rfloor + a$, where $0 \leq a < q$. Without any loss of generality, suppose $\tau_0 \geq \tau_1 \geq ... \geq \tau_{q-1}$. Then since $r = \sum_{i=0}^{q-1} \tau_i$,

$$\tau_0 = \tau_1 = \cdots = \tau_{a-1} = \lceil \frac{r}{q} \rceil,$$

$$\tau_a = \tau_{a+1} = \cdots = \tau_{q-1} = \lfloor \frac{r}{q} \rfloor.$$

Thus $[\tau_0, \tau_1, ..., \tau_{q-1}]$ is a fair partition of r since

$$r = \underbrace{\lceil \frac{r}{q} \rceil + ... + \lceil \frac{r}{q} \rceil}_{a} + \underbrace{\lfloor \frac{r}{q} \rfloor + ... + \lfloor \frac{r}{q} \rfloor}_{q-a}.$$

From (3) we have the following inequality

$$2(\binom{a}{2}\lceil\frac{r}{q}\rceil^2+\binom{q-a}{2}\lfloor\frac{r}{q}\rfloor^2+a(q-a)\lfloor\frac{r}{q}\rfloor\lceil\frac{r}{q}\rceil)\geq\rho(n-1) \quad (4)$$

Notice that if $a=0$ meaning that q divides r, then inequality (4) coincides with the inequality given by Levenshtein in [6]:

$$r\geq\sqrt{\frac{q\rho(n-1)}{q-1}}. \quad (5)$$

Otherwise if $a>0$, replacing a and $\lceil\frac{r}{q}\rceil$ with $r-q\lfloor\frac{r}{q}\rfloor$ and $\lfloor\frac{r}{q}\rfloor+1$ in (4), respectively, then

$$r(r-1)+(q-2r)\lfloor\frac{r}{q}\rfloor+q\lfloor\frac{r}{q}\rfloor^2\geq\rho(n-1). \quad (6)$$

Furthermore, suppose that q divides r then inequality (5) can be obtained from inequality (6) by replacing $\lfloor\frac{r}{q}\rfloor$ with r/q. Thus, inequality (6) holds for both situations. The sufficient and necessary conditions of equality follow directly from the proof. ∎

Theorem 2.5 *If* $\sqrt{\frac{q\rho(n-1)}{q-1}}$ *is a square-free integer, then*

$$r_q(n,\rho)\geq\sqrt{\frac{q\rho(n-1)}{q-1}}+1.$$

Otherwise, $r_q(n,\rho)\geq\sqrt{\frac{q\rho(n-1)}{q-1}}$.

Proof: First, we will show that if parameters n, q, and ρ satisfy the following conditions, then $r_q(n,\rho)\geq\sqrt{\frac{q\rho(n-1)}{q-1}}+1$.

$$\sqrt{\frac{q\rho(n-1)}{q-1}} \text{ is an integer,}$$

$$\text{and} \quad \sqrt{\frac{\rho(n-1)}{q(q-1)}} \text{ is not an integer.}$$

Second we will prove that if $\sqrt{\frac{q\rho(n-1)}{q-1}}$ is a square-free integer, then $\sqrt{\frac{\rho(n-1)}{q(q-1)}}$ can not be integral.

As we mentioned in the argument of lemma 2.4, if q divides r then our inequality (6) coincides with Levenshtein's (5). Thus let's take a look at the case when $\sqrt{\frac{q\rho(n-1)}{q-1}}$ is an integer but $\sqrt{\frac{\rho(n-1)}{q(q-1)}}$ is not. Then

$$\begin{aligned}\rho(n-1) &= \sqrt{\tfrac{q\rho(n-1)}{q-1}}(\sqrt{\tfrac{q\rho(n-1)}{q-1}}-1)+(q-2\sqrt{\tfrac{q\rho(n-1)}{q-1}})(\tfrac{\sqrt{\frac{q\rho(n-1)}{q-1}}}{q}) \\ &\quad +q(\tfrac{\sqrt{\frac{q\rho(n-1)}{q-1}}}{q})^2 \\ &> \sqrt{\tfrac{q\rho(n-1)}{q-1}}(\sqrt{\tfrac{q\rho(n-1)}{q-1}}-1)+(q-2\sqrt{\tfrac{q\rho(n-1)}{q-1}})\lfloor\tfrac{\sqrt{\frac{q\rho(n-1)}{q-1}}}{q}\rfloor \\ &\quad +q(\lfloor\tfrac{\sqrt{\frac{q\rho(n-1)}{q-1}}}{q}\rfloor)^2.\end{aligned}$$

This implies that $\sqrt{\frac{q\rho(n-1)}{q-1}}$ does not satisfy inequality (6). Hence $r \geq \sqrt{\frac{q\rho(n-1)}{q-1}}+1$.

Suppose that $\sqrt{\frac{q\rho(n-1)}{q-1}}$ is a square-free integer. Then its square is an integer and there is some integer b such that

$$q\rho(n-1) = (q-1)b. \tag{7}$$

Since $q > 1$ by hypothesis, $\gcd(q, q-1) = 1$. Hence (7) implies $q|b$. Let $b = qc$ for some integer $c > 0$. Then from (7), $\rho(n-1) = (q-1)c$.

Now either $\sqrt{\frac{\rho(n-1)}{q(q-1)}}$ is an integer or it is not. We show a contradiction by assuming it is an integer. If $\sqrt{\frac{\rho(n-1)}{q(q-1)}}$ is an integer, then its square is integral as well. Thus

$$\rho(n-1) = q(q-1)d \tag{8}$$

for some integer $d > 0$. Therefore, $(q-1)c = \rho(n-1) = q(q-1)d$ or $c = qd$. Hence there exists an integer $d > 0$ with $b = qc = q^2d$. It would imply that $\sqrt{\frac{q\rho(n-1)}{q-1}}$ has a repeated factor. ∎

When $\sqrt{\frac{q\rho(n-1)}{q-1}}$ is an integer, let

$$\frac{q\rho(n-1)}{q-1} = p_1^{2k_1}p_2^{2k_2}\dots p_s^{2k_s}$$

where p_i's are distinct primes and k_i's are positive integers. Since q and $q-1$ are relative prime, q can not have any other prime factor except for p_i's. Suppose $q = p_1^{t_1}\dots p_s^{t_s}$ where $0 \leq t_i \leq 2k_i$, then

$$\sqrt{\frac{\rho(n-1)}{q(q-1)}} = p_1^{k_1-t_1}\dots p_s^{k_s-t_s}.$$

When $\sqrt{\frac{\rho(n-1)}{q(q-1)}}$ is not an integer, there exists some i such that $t_i > k_i$, i.e. q has repeated prime factors and $\mu(q) = 0$. Hence the minimal value of q is 4 when equality in our inequality holds but Levenshtein's inequality does not give the actual lower bound. One can verify the following two infinity sequences reach our lower bound in stead of Levenshtein's: $q = 4$, $\rho = 3$, and $n = (2k+1)^2 + 1$; $q = 4$, $\rho = 1$ and $n = 3(2k+1)^2 + 1$ where $k \in \mathbf{Z}^+$.

Acknowledgments

This research was supported by NSF Grant CCR-0310832. The author is thankful to Professor Vladimir D. Tonchev for introducing this problem and helpful discussion and support. The author also thanks the referees for valuable comments.

References

[1] D. J. Clague, *New classes of synchronous codes*, IEEE Trans. on Electronic Computers **EC-16**, 290-298.

[2] H. C. Crick, J. S. Griffith, and L. E. Orgel, *Codes withour commas*, Proc. Nat. Acad. Sci. 43 (1957), 416-421.

[3] L. J. Cumming, *A Family of Circular Systematic Comma-Free Codes*, (to appear in this volume).

[4] S. W. Golomb, B. Gordon, L. R. Welch, *Comma-free code*, Canad. J. Math., vol. 10, no. 2, pp. 202-209, 1958.

[5] Y. Mutoh, and V. D. Tonchev, *Difference Systems of Sets and Cyclotomy*, Discrete Math., (to appear).

[6] V. I. Levenshtein, *One method of constructing quasilinear codes providing synchronization in the presence of errors*, Problems of Information Transmission, vol. 7, No. 3 (1971), 215-222.

[7] V. I. Levenshtein, *Combinatorial problems motivated by comma-free codes*, J. of Combinatorial Designs, vol. 12, issue 3 (2004), 184-196.

[8] V. D. Tonchev, *Difference systems of sets and code synchronization*, Rendiconti del Seminario Matematico di Messina , Series II, vol. 9 (2003), 217-226.

[9] V. D. Tonchev, *Partition of difference sets and code synchronization*, Finite Fields and their Applications, (to appear).