21^{th} Nearring And Nearfield Conference
26 July–1 August, 2009
Vorau, Styria

The Cardinality of Some Symmetric Differences

Po-Yi Huang, Wen-Fong Ke and Günter F. Pilz

July 27, 2009

Linear codes

- F: a finite field.
- V: an n dimensional vector space over F.
- B: a fixed ordered basis, for convenience, take the standard basis.
- Any subspace $\mathcal{C}$ of V is a linear code.
- The (Hamming) distance $d(v_1, v_2)$ of two vectors $v_1 = (a_1, \ldots, a_n)$ and $v_2 = (b_1, \ldots, b_n)$ in $\mathcal{C}$ is the number of i's such that $a_i \neq b_i$.
- The minimal distance $d = d_{\mathcal{C}}$ is $\min\{d(v_1, v_2) \mid v_1, v_2 \in \mathcal{C} \text{ and } v_1 \neq v_2\}$.
- The weight $\mathrm{wt}(v)$ of $v = (a_1, \ldots, a_n) \in V$ is the number of i's with $a_i \neq 0$.
- $d_{\mathcal{C}} = \min\{\mathrm{wt}(v) \mid v \in \mathcal{C},\ v \neq 0\}$.

Linear codes

A linear code with minimal distance d can correct up to $\lfloor \frac{d-1}{2} \rfloor$ errors, and there is a standard way of doing it.

- The goal of coding theory is to find codes in vector spaces V of dimension n
  - with smaller n yet large minimal distance d, and
  - with some easy ways of encoding and decoding.

Multiplication codes (Cyclic codes)

- $f = x^n - 1$ in $F[x]$ and $J = (f)$ the ideal generated by f.
- $V = F[x]/J$, which is a principal ideal ring as well a vector space of dimension n over F.
- Take any nonzero $g \in F[x]$ of degree $n - m$.
- The the ideal $\mathcal{C}$ in V generated by $\overline{g} = g + J$ is a subspace of V of dimension m.
- A word $(a_0, \ldots, a_{m-1}) \in F^m$ is identified as the polynomial $a_0 + a_1 x + \ldots a_{m-1} x^{m-1}$ of degree at most $m - 1$.
- A polynomial $h \in F[x]$ of degree at most $m - 1$ is encoded as $h \cdot g + J$ in V. This makes $\mathcal{C}$ a *multiplication code*.
- Given a polynomial $k \in F[x]$ of degree at most $n - 1$. Then $k + J$ is in $\mathcal{C}$ if and only if $k \equiv g \cdot h \pmod{x^n - 1}$.

Multiplication codes (Cyclic codes)

- Take $h = a_0 + a_1x + \cdots + a_{n-1}x^{n-1} \in F[x]$ such that $\overline{h(x)} \in \mathcal{C}$. Then $\overline{x} \cdot \overline{h}$ since $\mathcal{C}$ is an ideal.
- Since $x^n \equiv 1$ in V,
  $xh \equiv a_{n-1} + a_0x + \cdots + a_{n-2}x^{n-1} \pmod{x^n - 1}$.
  Thus $a_{n-1} + a_0x + \cdots + a_{n-2}x^{n-1} + J$ is a codeword as well.
  This makes $\mathcal{C}$ a cyclic code.
- Cyclic codes are easy in computation. With suitable choices of g, one can get better decoding algorithm.
- Many important codes are binary multiplication codes (e.g. BCH codes and Reed-Solomon codes).
- The Reed-Solomon code $RS(2^r, d)$ uses g of degree $d - 1$, and has a minimal distance d.

Composition codes

- Let $f = x + x^2 + \cdots + x^k \in \mathbb{Z}_2[x]$. Let $m \geq 2$ and $n = km$.
- Let $\mathcal{C} = C(f, m)$ be the subspace of $V = \{a_1 x + \cdots + a_n x^n \mid a_i \in \mathbb{Z}_2\}$ generated by $f \circ x$, $f \circ x^2$, $\ldots$, $f \circ x^m$.
- A word $(a_1, \ldots, a_m) \in \mathbb{Z}_2^m$ is encoded as

$$\begin{aligned} & a_1 f \circ x + a_2 f \circ x^2 + \cdots + a_m f \circ x^m \\ = & a_1(x + x^2 + \cdots + x^r) + a_2(x^2 + x^4 + \cdots + x^{2k}) + \ldots \\ & + a_m(x^m + x^{2m} + \cdots + x^{mk}) \in V. \end{aligned}$$

- $\mathcal{C}$ is referred as a composition code, and a general theory has been studied by Fuchs (1992).
- The minimal distance $d_{\mathcal{C}}$ is k if $k \leq 6$ [Pilz 1992].

A general question

- What is the minimal distance $d_{\mathcal{C}}$ in general? Equivalently, what is the minimal weight of nonzero codewords?
  The answer is not known except for $k \leq 6$.

- For a codeword $v = a_1 f \circ x + a_2 f \circ x^2 + \cdots + a_m f \circ x^m \in \mathcal{C}$, $\mathrm{wt}(v)$ is the cardinality of the symmetric differences

$$\{x^{i_1}, \ldots, x^{i_1 k}\} \bigtriangleup \{x^{i_2}, \ldots, x^{i_2 k}\} \bigtriangleup \cdots \bigtriangleup \{x^{i_s}, \ldots, x^{i_s k}\}$$

  where $\{i_1, \ldots, i_s\} = \{i \mid a_i \neq 0\}$.
- For example, if $f = x + x^2 + x^3$ and $v = f \circ x + f \circ x^3$, then $\mathrm{wt}(v) = 4$:

$$\begin{aligned} v &= (x + x^2 + x^3) \circ x + (x + x^2 + x^3) \circ x^3 \\ &= x + x^2 + x^6 + x^9. \end{aligned}$$

- $\{x, x^2, x^3\} \bigtriangleup \{x^3, x^6, x^9\} = \{x, x^2, x^6, x^9\}$.

The 1-2-3 Conjecture

- A special situation: what is the weight of the codeword $f \circ x + f \circ x^2 + \cdots + f \circ x^m$? That is, what is the cardinality of

$$\{x^1, \ldots, x^k\} \bigtriangleup \{x^2, \ldots, x^{2k}\} \bigtriangleup \cdots \bigtriangleup \{x^m, \ldots, x^{mk}\}?$$

The 1-2-3 Conjecture

For all m, the cardinality of the symmetric differences

$$\{x^1, \ldots, x^k\} \bigtriangleup \{x^2, \ldots, x^{2k}\} \bigtriangleup \cdots \bigtriangleup \{x^m, \ldots, x^{mk}\}$$

is at less k.

- True for $k = 7$ and $k = 8$ [E. Fried (Budapest)].
- True for $k \geq 10^{12}$ [P. Fuchs (Linz)].

The 1-2-3 Conjecture

The Restricted 1-2-3 Conjecture

For $k \leq m$, the cardinality of the symmetric differences

$$\{x^1, \ldots, x^k\} \vartriangle \{x^2, \ldots, x^{2k}\} \vartriangle \cdots \vartriangle \{x^m, \ldots, x^{mk}\}$$

is at less k.

The Theorem

Theorem (Huang, Pilz, K)

For $k \leq m$, the cardinality of the symmetric differences

$$\{1, \ldots, k\} \triangle \{2, \ldots, 2k\} \triangle \cdots \triangle \{m, \ldots, mk\}$$

is at less m.

Notation for the proof

- $I_k := \{1, 2, \ldots, k\}$.
- For $s \in \mathbb{N}$, $sI_k := \{s, 2s, \ldots, ks\}$.
- For $1 \leq u < v$,

$$D_{k\times[u,v]} := uI_k \bigtriangleup (u+1)I_k \bigtriangleup \ldots \bigtriangleup vI_k.$$

- $D_{k\times v} := D_{k\times[1,v]}$ and $d_k(v) = |D_{k\times v}|$.
- If $1 < s < v$, then

$$D_{k\times v} = D_{k\times s} \bigtriangleup D_{k\times[s+1,v]}.$$

- $D_{k\times v} = D_{v\times k}$ for all k and v and $|D_{k\times k}| = k$.

Case 1. $k < m = k + w \leq 2k$

Assume that $k < m = k + w \leq 2k$.

- The goal is to find that $|D_{k\times[k+1,k+w]}| \geq 3k$.
- If $\ell = \gcd(s,t)$, then $|sI_k \cap tI_k| \leq \ell - 1$.
  The number of cancelations taking place in $sI_k \triangle tI_k$ is at most $2(\ell - 1)$.
- There are at most $\lceil \frac{w}{\ell} \rceil$ a's with $\ell \mid a$ and $k + 1 \leq a \leq k + w$.
- The total number of cancelations occurring in $(k+1)I_k \triangle \ldots \triangle (k+w)I_k$ is at most $\sum_{\ell=2}^{w-1} \binom{\lceil \frac{w}{\ell} \rceil}{2} \cdot 2(\ell - 1) < 2w^2 \cdot \ln(w-1)$.
- There are at least $kw - 2w^2 \ln(w-1)$ elements in $D_{k\times[k+1,k+w]}$.

Case 1. $k < m = k + w \leq 2k$

- Now, $|D_{k\times(k+w)}| \geq 2k \Leftarrow kw - 2w^2 \cdot \ln(w-1) \geq 3k \Leftrightarrow$ $k \geq \frac{2w^2 \cdot \ln(w-1)}{w-3}$.
- For each given k,
  set $w_k = \max\{w \geq 1 \mid w \text{ satisfies } k \geq \frac{2w^2 \cdot \ln(w-1)}{w-3}\}$.
- $f(x) = \frac{2x^2 \cdot \ln(x-1)}{x-3}$ is increasing for $x \geq 5$.

Lemma

Suppose that $w_k \geq 5$ and that there are two distinct primes among $k+1, \ldots, k+w_k$. Then $D_{k\times(k+w)}$ has at least $2k$ elements for all w with $5 \leq w \leq k$.

Case 1. $k < m = k + w \leq 2k$

- A *prime gap* is the difference between two successive prime numbers.
- One writes $g(p)$ for the the gap $q - p$, where q is the next prime to p. E.g. $g(11) = 13 - 11 = 2$.
- A prime gap is *maximal* if it is larger than all gaps between smaller primes. The n-th maximal prime gap is denoted by g_n.
- For example, $g_1 = 1$, $g_2 = 2$, $g_3 = 4$, $g_4 = 6$, and $g_{11} = 9551$. Thus, for any prime $p < 9551$, the prime gap $g(p)$ is less than 36, and so there must be a prime in the set $\{p + 1, \ldots, p + 36\}$.
- If p is prime, $p > 2k$, and $g_t < \frac{w_k}{2}$ a maximal prime gap, q the smallest prime with $g(q) = g_t$ and $p \leq q$, then there exist at least two primes in $\{k + 1, k + 2, \ldots, k + w_k\}$.

Case 1. $k < m = k + w \leq 2k$

Combining w_k and g_n, we could argue that

- At least two primes exists between $k+1$ and $k+w_k$ if $k > 70919$.
- Any prime $p \leq 70919$ has $g(p) \leq 72$, while for $2000 < k \leq 70919$, $w_k \geq 189 > 2g(p)$.
- Any prime $p \leq 2000$ has $g(p) \leq 34$, while for $600 < k \leq 2000$, $w_k \geq 68 \geq 2g(p)$.
- Any prime $p \leq 600$ has $g(p) \leq 18$, while for $300 < k \leq 600$, $w_k \geq 38 > 2g(p)$.
- For $k < 300$ and $k < m \leq 2k$, a simple computer check shows that $d_k(m) \geq m$.

Case 2. $2k < m$

- It suffices to assume that $m \leq \mathrm{LCM}(I_k)$.
- Set $\mathcal{P} = \{p \mid p \text{ is a prime and } \max\{k, \sqrt{m}\} < p \leq m\}$.
- If $p \in \mathcal{P}$, then $\lfloor \frac{m}{p} \rfloor < m$, and

$$pI_k \vartriangle 2pI_k \vartriangle \ldots \vartriangle \lfloor \tfrac{m}{p} \rfloor pI_k = p(I_k \vartriangle 2I_k \vartriangle \ldots \vartriangle \lfloor \tfrac{m}{p} \rfloor I_k)$$

  which has at less $\max\{k, \lfloor \frac{m}{p} \rfloor\}$ many elements (induction on m with $m = 2k$ as the base).
- If $p, q \in \mathcal{P}$ are distinct, then $(sp)I_k \cap (tq)I_k = \varnothing$ for any $1 \leq s \leq \lfloor \frac{m}{p} \rfloor$ and $1 \leq t \leq \lfloor \frac{m}{q} \rfloor$. Thus,

  Each $p \in \mathcal{P}$ contributes at least $\max\{k, \lfloor \frac{m}{p} \rfloor\}$ elements.

- Goal: Show that $\sum_{p \in \mathcal{P}} \max\{k, \lfloor \frac{m}{p} \rfloor\} \geq m$.

Case 2-1. $2k < m \leq k^2$

Assume $2k < m \leq k^2$.

- $\max(k, \sqrt{m}) = k$.
- $\mathcal{P} = \{p \mid p \text{ is a prime and } k < p \leq m\}$.
- $\sum_{p \in \mathcal{P}} \max\{k, \lfloor \frac{m}{p} \rfloor\} = |\mathcal{P}| \cdot k$.
- Set $\lceil \frac{m}{k} \rceil = n$. Then, $n \geq 3$ and $kn \geq m$.
- Just have to show that $|\mathcal{P}| \geq n$.

- If $k \geq 21$, then $|\mathcal{P}| = \pi(m) - \pi(k) \geq n$.
- For $1 \leq k \leq 20$, and $2k < m \leq k^2$, we use computer to verify.

Case 2-2. $k^2 < m$

Assume that $k^2 < m$.

- $\sum_{p \in \mathcal{P}} \max\{k, \lfloor \frac{m}{p} \rfloor\} \geq \frac{m}{2} + \frac{m}{\ln m} \cdot (k - \ln(k+1) - 2.52)$.
- $\frac{k - \ln(k+1) - 2.52}{\ln m} \geq \frac{1}{2}$ for $k \geq 8$.
- For $k \leq 7$, and $k^2 < m \leq \mathrm{LCM}(I_k)$, use computer to verify.

Some results used from number theory:

Theorem

1. $2^k \leq \mathrm{LCM}(I_k) \leq 4^k$.
2. $\pi(x) > x/\ln x$ *for* $x \geq 17$.
3. $\pi(x) < 1.25506x/\ln x$ *for* $x > 1$.
4. $\pi(2x) - \pi(x) > 3x/(5 \ln x)$ *for* $x > 20.5$.

BEWARE: THE BEAST IS STILL OUT THERE.

Go find the minimal distances of the binary composition codes $C(x + x^2 + \cdots + x^k, m)$ for $k \geq 7$.