Large Constant Dimension Codes and Lexicodes

Natalia Silberstein Tuvi Etzion
Department of Computer Science
Technion-Israel Institute of Technology

ALCOMA 10 Germany

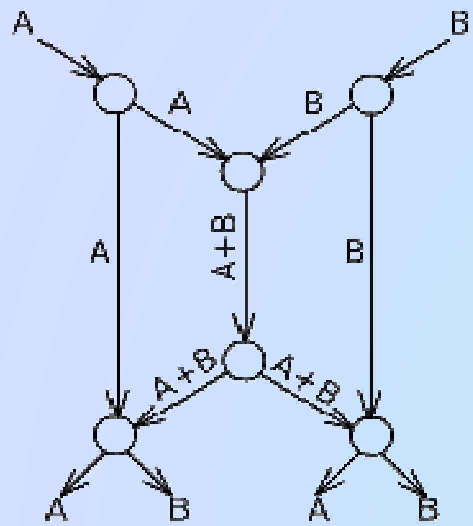
Definitions

- Let $\mathbb{F}_q$ be a finite field of size q
- The Grassmannian space, $\mathcal{G}_q(n,k)$, is the set of all k-dimensional subspaces of $\mathbb{F}_q^n$
- $\mathcal{G}_q(n,k)$ is a metric space with the distance function

  $d_S(X,Y) = \dim(X) + \dim(Y) - 2\dim(X \cap Y)$
- A $\mathbb{C} \subseteq \mathcal{G}_q(n,k)$ is an $(n,M,d,k)_q$ constant dimension code if $|\mathbb{C}| = M$, and $d_S(X,Y) \geq d$ for all $X \neq Y \in \mathbb{C}$

Motivation

- Koetter and Kschischang (2007) showed an application of error-correcting codes in $\mathcal{G}_q(n,k)$ to random network coding



Our goal:

Construction of large constant dimension codes

Lexicodes

- Lexicographic codes (lexicodes) are greedily generated codes
- The construction of a lexicode with a minimum distance d :
  - starts with the set $S = \{S_0\}$, where S_0 is the first element in a lexicographic order;
  - greedily adds the lexicographically first element whose distance from all the elements of S is at least d.

Outline

- Representation of subspaces
- Multilevel structure of constant dimension codes
- Search method for constant dimension lexicodes
- Lexicodes with a seed
- Conclusion and open problems

Representation of subspaces

- A subspace $X \in \mathcal{G}_q(n,k)$ can be represented by the k x n generator matrix RE(X) in reduced row echelon form

Example. Let X = {(0000000), (1011000), (1001101), (1010011), (<u>0010101</u>), (<u>0001011</u>), (0011110), (<u>1000110</u>)} be in $\mathcal{G}_2(7,3)$. Then

$$\mathrm{RE}(X) = \begin{pmatrix} 1000110 \\ 0010101 \\ 0001011 \end{pmatrix}$$

Identifying vectors

- For each subspace $X \in \mathcal{G}_q(n,k)$ there is an identifying vector $v(X) \in \{0,1\}^n$ of weight k
  - The ones in v(X) are in the positions where RE(X) has leading ones

Example. If $X \in \mathcal{G}_2(7,3)$ is given by

$$\mathrm{RE}(X) = \begin{pmatrix} 1000110 \\ 0010101 \\ 0001011 \end{pmatrix} \text{ then}$$

$v(X)$=1011000

Identifying vectors

- For each subspace $X \in \mathcal{G}_q(n,k)$ there is an identifying vector $v(X) \in \{0,1\}^n$ of weight k
  - The ones in v(X) are in the positions where RE(X) has leading ones

Example. If $X \in \mathcal{G}_2(7,3)$ is given by

$$\mathrm{RE}(X) = \begin{pmatrix} 1\bullet 00 \bullet\bullet\bullet \\ 0010 \bullet\bullet\bullet \\ 0001 \bullet\bullet\bullet \end{pmatrix} \text{ then}$$

$$v(X) = 1011000$$

Echelon Ferrers Form

- For each vector $v \in \{0,1\}^n$ of weight k there is the echelon Ferrers form, EF(v):

Example. Let $v = 1\,011000$ then

$$\mathrm{EF}(v) = \begin{pmatrix} 1\bullet 00\bullet\bullet\bullet \\ 0010\bullet\bullet\bullet \\ 0001\bullet\bullet\bullet \end{pmatrix}$$

Echelon Ferrers Form

- For each vector $v \in \{0,1\}^n$ of weight k there is the echelon Ferrers form, EF(v):

Example. Let $v = 1011000$ then

$$\mathrm{EF}(v) = \begin{pmatrix} 1 & \bullet & 0 & 0 & \bullet & \bullet & \bullet \\ 0 & 0 & 1 & 0 & \bullet & \bullet & \bullet \\ 0 & 0 & 0 & 1 & \bullet & \bullet & \bullet \end{pmatrix}, \quad \mathcal{F} = \begin{array}{cccc} \bullet & \bullet & \bullet & \bullet \\ & \bullet & \bullet & \bullet \\ & \bullet & \bullet & \bullet \end{array}$$

- The dots of EF(v) form Ferrers diagram, $\mathcal{F}$.

Echelon Ferrers Form

- For each vector $v \in \{0,1\}^n$ of weight k there is the echelon Ferrers form, EF(v):

Example. Let $v = 1\,011000$ then

$$EF(v) = \begin{pmatrix} 1 & \bullet & 0 & 0 & \bullet & \bullet & \bullet \\ 0 & 0 & 1 & 0 & \bullet & \bullet & \bullet \\ 0 & 0 & 0 & 1 & \bullet & \bullet & \bullet \end{pmatrix}, \quad \mathcal{F} = \begin{matrix} \bullet & \bullet & \bullet & \bullet \\ & \bullet & \bullet & \bullet \\ & \bullet & \bullet & \bullet \end{matrix}, \quad RE(X) = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

- The dots of EF(v) form Ferrers diagram, $\mathcal{F}$.
- If we substitute some elements of $\mathbb{F}_q$ in the dots of EF(v), we obtain RE(X) for some X $\in \mathcal{G}_q(n,k)$

Multilevel Structure of $\mathcal{G}_q(n,k)$

- All the binary vectors of the length n and weight k can be considered as the identifying vectors of all the subspaces in $\mathcal{G}_q(n,k)$
- These $\binom{n}{k}$ vectors partition $\mathcal{G}_q(n,k)$ into the $\binom{n}{k}$ different classes, which are called Schubert cells.
- Each Schubert cell contains all the subspaces with the same given echelon Ferrers form.

Multilevel Structure of Constant Dimension Codes

- We partition all the codewords of a constant dimension code into different classes (sub-codes), by the identifying vectors.
- First level: the set of identifying vectors.
- Second level: subspaces corresponding to these vectors.

Multilevel Structure of Constant Dimension Codes

- Let $\mathbb{C} \subseteq \mathcal{G}_q(n,k)$,
- Let $\{v_1,\ldots, v_t\}$ be all the different identifying vectors in $\mathbb{C}$.
- Then $\{\mathbb{C}_1,\ldots,\mathbb{C}_t\}$ is the partition of $\mathbb{C}$ into t sub-codes, where $v(\mathbb{C}_i) = v_i$, for each $X \in \mathbb{C}_i$.

<u>Example.</u> Let $n = 6$, $k = 3$, $q = 2$, $d = 4$, $a_i \in \mathbb{F}_2$

v_1=111000 $\quad$ v_2=100110 $\quad$ v_3=010101 $\quad$ v_4=001011

$$\begin{pmatrix} 100\bullet\bullet\bullet \\ 010\bullet\bullet\bullet \\ 001\bullet\bullet\bullet \end{pmatrix} \quad \begin{pmatrix} 1\bullet\bullet 00\bullet \\ 00010\bullet \\ 00001\bullet \end{pmatrix} \quad \begin{pmatrix} 01\bullet 0\bullet 0 \\ 0001\bullet 0 \\ 000001 \end{pmatrix} \quad \begin{pmatrix} 001\bullet 00 \\ 000010 \\ 000001 \end{pmatrix}$$

$$\mathbb{C}_1 = \left\{ \begin{pmatrix} 100\,a_1a_2a_3 \\ 010\,a_4a_5a_6 \\ 001\,a_7a_8a_9 \end{pmatrix} \right\} \quad \mathbb{C}_2 = \left\{ \begin{pmatrix} 1a_1a_2 00a_3 \\ 0\,0\,0\,10a_4 \\ 0\,0\,0\,01a_5 \end{pmatrix} \right\} \quad \mathbb{C}_3 = \left\{ \begin{pmatrix} 01a_1 0a_2 0 \\ 000\,1a_3 1 \\ 000\,0\,0\,1 \end{pmatrix} \right\} \quad \mathbb{C}_4 = \left\{ \begin{pmatrix} 001a_1\,00 \\ 000\,0\,10 \\ 000\,0\,01 \end{pmatrix} \right\}$$

Identifying Vectors and Subspace Distance

- Theorem 1. $d_S(X, Y) = d_H(v(X), v(Y)) + 2\mathrm{rank}(Z_{xy})$

- Corollary 1. $d_S(X, Y) \geq d_H(v(X), v(Y))$

- Corollary 2. If $v(X) = v(Y)$, then

  $d_S(X, Y) = 2\mathrm{rank}(RE(X) - RE(Y))$

Search Method for Constant Dimension Lexicodes

- In each step we have the current code $\mathbb{C}$ and the set of subspaces not examined yet.
- Order the set of all binary words of length n and weight k

  (they are the candidates to be the identifying vectors of codewords)
- For each candidate for an identifying vector v search for a sub-code $\mathbb{C}_v$:
  - For each next subspace X calculate the distance between X and $\mathbb{C}$, and add X to $\mathbb{C}$ if this distance at least d.

Search Method for Constant Dimension Lexicodes

- In each step we have the current code $\mathbb{C}$ and the set of subspaces not examined yet.
- Order the set of all binary words of length n and weight k

  (they are the candidates to be the identifying vectors of codewords)
- For each candidate for an identifying vector v search for a sub-code $\mathbb{C}_v$:
  - For each next subspace X calculate the distance between X and $\mathbb{C}$, and add X to $\mathbb{C}$ if this
  - Optimization:
    - first calculate the Hamming distance between the identifying vectors of representatives of sub-codes to determine a lower bound on the subspace distance,
    - only if necessary, calculate the subspace distance

Corollary 1: $d_S(Y, Z) \geq d_H(v(Y), v(Z))$

Constant Dimension Codes and Rank-Metric Codes

- Let $X \in \mathcal{G}_q(n,k)$
- R(X) is the k x $(n\text{-}k)$ sub-matrix of RE(X) with the columns indexed by zeroes of v(X)
- $d_S(X,Y) = 2\, d_R(R(X),R(Y))$, for all $X,Y \in \mathbb{C}_i$

  where $d_R(A,B)=\text{rank}(A-B)$, for any two matrices A and B of the same size
- Let $\mathcal{F}$ be a Ferrers diagram with m rows and η columns
- A code C is an $[\mathcal{F}, \rho, \delta]$ Ferrers diagram rank-metric code if
  - it forms a linear subspace of $\mathbb{F}_q^{m \times \eta}$ of dimension ρ ;
  - for each $A \neq B \in C$, $d_R(A,B) \geq \delta$
  - each codeword has zeroes in all entries not in $\mathcal{F}$

Upper Bound on Size of Ferrers Diagram Rank-Metric Codes

- Let dim($\mathcal{F}$, δ) be the largest possible dimension of an [$\mathcal{F}$,ρ,δ] code.
- Theorem 2. dim($\mathcal{F}$, δ) ≤ min $\{v_i\}$, where v_i, $0 \leq i \leq \delta\text{-}1$, is the number of dots in $\mathcal{F}$ which are not contained in the first i rows or the rightmost $\delta\text{-}1\text{-}i$ columns
- It is not known whether this upper bound is attained for all parameters.
- A code which attains this bound is called maximum rank distance Ferrers diagram code (MRD code).

Properties of Constant Dimension Codes

- For each $\mathbb{C}_i \subseteq \mathbb{C}$ define a Ferrers diagram rank-metric code

$$R(\mathbb{C}_i) = \{R(X) : X \in \mathbb{C}_i\}$$

- $R(\mathbb{C}_i)$ will be called unlifted code of $\mathbb{C}_i$
- $d_S(\mathbb{C}_i, \mathbb{C}_j) = \min\{d_S(X, Y) : X \in \mathbb{C}_i, Y \in \mathbb{C}_j\}$
- $d_S(\mathbb{C}_i, \mathbb{C}_j) \geq d_H(v_i, v_j)$
- Lemma 1. Let $\mathbb{C}_i$, $\mathbb{C}_j$ be two sub-codes of $\mathbb{C} \subseteq \mathcal{G}_q(n,k)$, such that $X \in \mathbb{C}_i$, $Y \in \mathbb{C}_j$ and RE(X) and RE(Y) are some column permutation of the matrix $(I_k 0_{k\times(n-k)})$. Then

$$d_S(\mathbb{C}_i, \mathbb{C}_j) = d_H(v_i, v_j)$$

- Corollary 3. Let $\mathbb{C}$ be an $(n,M,d,k)_q$ code. If $d_H(v_i, v_j) < d$ then at least one unlifted code ($R(\mathbb{C}_i)$ or $R(\mathbb{C}_j)$) is nonlinear.

Multilevel Construction for an $(n,M,d=2\delta,k)_q$ code $\mathbb{C}^{\mathrm{ML}}$

- First level. Take a binary constant weight code C of length n, weight k and minimum distance δ to be the set of identifying vectors of $\mathbb{C}^{\mathrm{ML}}$

- Second level. For each constant weight codeword $v_i \in C$ construct a sub-code $\mathbb{C}_i$ such that R($\mathbb{C}_i$) is a Ferrers diagram MRD code with the minimum distance δ.

Example: $(8,4605,4,4)_2$ lexicode $\mathbb{C}^{lex}$ vs. $(8,4573,4,4)_2$ code $\mathbb{C}^{ML}$

★ - nonlinear unlifted code (coset of linear code)

i	id.vector v_i	size of $\mathbb{C}_i^{lex}$	size of $\mathbb{C}_i^{ML}$
1	11110000	4096	4096
2	11001100	256	256
3	10101010	64	64
4	10011010	16 ★	–
5	10100110	16 ★	–
6	00111100	16	16
7	01011010	16 ★	16
8	01100110	16 ★	16
9	10010110	16	16
10	01101001	32	32
11	10011001	16 ★	16
12	10100101	16 ★	16
13	11000011	16	16
14	01010101	8	8
15	00110011	4	4
16	00001111	1	1

Properties of Constant Dimension Codes

- Lemma 2. Let $\mathbb{C}$ be an $(n,M,d,k)_q$ code. Let $\mathbb{C}_1 \subseteq \mathbb{C}$ be a sub-code with identifying vector $v_1 = \underbrace{11\ldots1}_{k}\underbrace{00\ldots0}_{n-k}$, such that R($\mathbb{C}_1$) is an MRD code. Then there is no codeword Y in $\mathbb{C}$ such that $d_H(v(Y),v_1) < d$.

- Corollary 4. If an $(n,M,d,k)_q$ code $\mathbb{C}$ contains a sub-code $\mathbb{C}_1$ such that R($\mathbb{C}_1$) is an MRD code, then the second sizewise Ferrers diagram of $\mathbb{C}$ corresponds to the identifying vector $v_2 = \underbrace{11\ldots1}_{k-\delta}\underbrace{00\ldots0}_{\delta}\underbrace{11\ldots1}_{\delta}\underbrace{00\ldots0}_{n-k-\delta}$.

Lexicodes with a Seed

- First step. Construct the maximal sub-code $\mathbb{C}_1$ which corresponds to the identifying vector $v_1 = \underbrace{11\ldots1}_{k}\underbrace{00\ldots0}_{n\text{-}k}$.

  (take any known MRD code as a unlifted code R($\mathbb{C}_1$).)
- Second step. Construct a sub-code $\mathbb{C}_2$ which corresponds to the identifying vector $v_2 = \underbrace{11\ldots1}_{k\text{-}\delta}\underbrace{00\ldots0}_{\delta}\underbrace{11\ldots1}_{\delta}\underbrace{00\ldots0}_{n\text{-}k\text{-}\delta}$

  (If there exists an MRD Ferrers diagram code, take any known construction of such code for R($\mathbb{C}_2$).)
- Third step. Construct the other sub-codes, according to the lexicode construction. (Examine only subspaces which are not pruned out by Lemma 2.)

Lexicodes with a Seed
(a variant)

- We can take as a seed any subset of codewords obtained by any given construction and to continue by applying the lexicode with a seed construction

Lexicodes with a Seed (Examples)

n	k	d	q	Size of lexicode with a seed	Size of previously known code
7	3	4	3	6691	6685
9	4	4	2	37649	36945
10	5	6	2	32890	32841

Conclusion and Open Problems

- We presented a search method for constant dimension codes based on their multilevel structure.
- Some of the codes obtained by this search are the largest known constant dimension codes
- Open Problems
  – Is the upper bound on the size of Ferrers diagram rank-metric codes is attainable for all parameters?
  – What is the best choice of identifying vectors for constant dimension codes?
  – Is there an optimal combination of linear Ferrers diagram rank-metric codes and cosets of linear codes to form a large constant dimension codes?

Thank you!