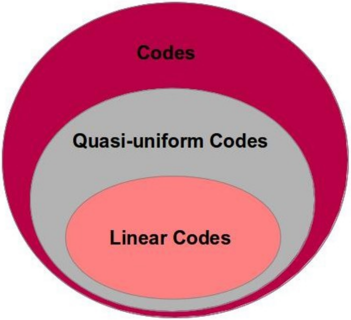
Group Theoretic Construction of Quasi-Uniform Codes

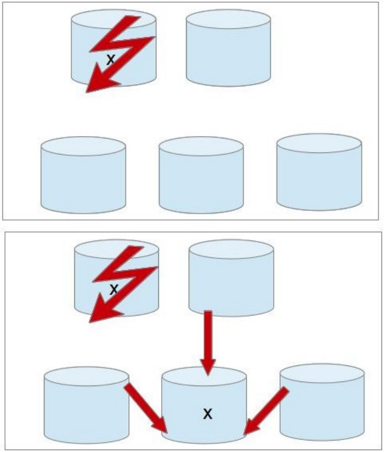
Eldho K Thomas
(Joint work with Frédérique Oggier)

Institute of Computer Science
University of Tartu
Estonia

Joint Estonian-Latvian Theory Days
Lilaste, October 15

What this talk is about





Introduction

- A code C of length n is a subset of $\mathcal{X}_1 \times \cdots \times \mathcal{X}_n$; $\mathcal{X}_i$ - alphabet for the i^{th} codeword symbol.

[1] T.H. Chan, A. Grant, T. Britz, "Properties of Quasi-Uniform Codes", 2010.

Introduction

- A code C of length n is a subset of $\mathcal{X}_1 \times \cdots \times \mathcal{X}_n$; $\mathcal{X}_i$ - alphabet for the i^{th} codeword symbol.
- Treat each codeword $(X_1, \ldots, X_n) \in C$ as a random vector with probability (for $\mathcal{N} = \{1, \ldots, n\}$)

$$P(X_{\mathcal{N}} = x_{\mathcal{N}}) = \begin{cases} 1/|C| & \text{if } x_{\mathcal{N}} \in C, \\ 0 & \text{otherwise.} \end{cases}$$

[1] T.H. Chan, A. Grant, T. Britz, "Properties of Quasi-Uniform Codes", 2010.

Introduction

- A code C of length n is a subset of $\mathcal{X}_1 \times \cdots \times \mathcal{X}_n$; $\mathcal{X}_i$ - alphabet for the i^{th} codeword symbol.
- Treat each codeword $(X_1, \ldots, X_n) \in C$ as a random vector with probability (for $\mathcal{N} = \{1, \ldots, n\}$)

$$P(X_{\mathcal{N}} = x_{\mathcal{N}}) = \begin{cases} 1/|C| & \text{if } x_{\mathcal{N}} \in C, \\ 0 & \text{otherwise.} \end{cases}$$

- A code C is quasi-uniform[1] if the induced codeword symbol random variables are uniformly distributed over their support.

[1] T.H. Chan, A. Grant, T. Britz, "Properties of Quasi-Uniform Codes", 2010.

Example

The [2,1] repetition code $\begin{array}{c|c} 0 & 0 \\ 1 & 1 \end{array}$ is quasi-uniform. The induced random variables X_1, X_2 take values $0, 1$ such that

$P(X_1 = 0) = P(X_2 = 0) = P(X_{12} = 00) = 1/2$.
But $P(X_{12} = 01) = P(X_{12} = 10) = 0$ since $01, 10 \notin \lambda(X_{12})$.

Example

The [2,1] repetition code $\begin{array}{c|c} 0 & 0 \\ 1 & 1 \end{array}$ is quasi-uniform. The induced random variables X_1, X_2 take values $0, 1$ such that

$P(X_1 = 0) = P(X_2 = 0) = P(X_{12} = 00) = 1/2$.
But $P(X_{12} = 01) = P(X_{12} = 10) = 0$ since $01, 10 \notin \lambda(X_{12})$.

Example

Consider the code $\begin{array}{c|c} 0 & 0 \\ 0 & 1 \\ 1 & 0 \end{array}$. Here $|\lambda(X_1)| = 2$ but the probabilities
$P(X_1 = 0) = 2/3, P(X_1 = 1) = 1/3$. Hence the code is not quasi-uniform.

Groups- Preliminaries

Definition

A group G is a set endowed with a binary operation satisfying:

1. *G is closed under the binary operation*
2. *The binary operation is associative*
3. *There exists an identity element 1 such that $1g = g1 = g$ for every $g \in G$*
4. *Every element is invertible*

Groups- Preliminaries

Definition

A group G is a set endowed with a binary operation satisfying:

1. *G is closed under the binary operation*
2. *The binary operation is associative*
3. *There exists an identity element 1 such that $1g = g1 = g$ for every $g \in G$*
4. *Every element is invertible*

Definition

Given a subgroup G_i of G, a **left coset** *(respectively right) of G_i in G is defined as*

$$gG_i = \{gh,\ h \in G_i\} \text{ respectively } G_i g = \{hg,\ h \in G_i\}.$$

Quasi-Uniform Codes from Groups

Theorem (Chan, Yeung, 2002)

[2] *For any finite group G and subgroups $G_1, \ldots, G_n$, $\exists$ n quasi-uniform discrete random variables $X_1, \ldots, X_n$ such that $\forall$ $\mathcal{A}$ of $\mathcal{N} = \{1, \ldots, n\}$, $P(X_{\mathcal{A}} = x_{\mathcal{A}}) = 1/[G : G_{\mathcal{A}}] = |G_{\mathcal{A}}|/|G|$; where $G_{\mathcal{A}} = \cap_{i \in \mathcal{A}} G_i$.*

X - a random variable uniformly distributed over G.

$X_i = XG_i$, the $[G : G_i]$ cosets of G_i.

[2] T. H. Chan and R. W. Yeung, "On a Relation Between Information Inequalities and Group Theory", 2002.

Quasi-Uniform Codes from Groups

Theorem (Chan, Yeung, 2002)

[2] *For any finite group G and subgroups $G_1, \ldots, G_n$, $\exists$ n quasi-uniform discrete random variables $X_1, \ldots, X_n$ such that $\forall$ $\mathcal{A}$ of $\mathcal{N} = \{1, \ldots, n\}$, $P(X_{\mathcal{A}} = x_{\mathcal{A}}) = 1/[G : G_{\mathcal{A}}] = |G_{\mathcal{A}}|/|G|$; where $G_{\mathcal{A}} = \cap_{i \in \mathcal{A}} G_i$.*

X - a random variable uniformly distributed over G.
$X_i = XG_i$, the $[G : G_i]$ cosets of G_i.

Corresponding quasi-uniform code can be obtained as follows:

	G_1	$\ldots$	G_n
$1 = g_1$	$g_1 G_1 = G_1$		$g_1 G_n = G_n$
g_2	$g_2 G_1$		$g_2 G_n$
$\vdots$	$\vdots$		$\vdots$
$g_{\|G\|}$	$g_{\|G\|} G_1$	$\ldots$	$g_{\|G\|} G_n$

[2] T. H. Chan and R. W. Yeung, "On a Relation Between Information Inequalities and Group Theory", 2002.

Example

$\mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_2$	G_1	G_2	G_3	G_4	G_5	G_6	G_7	G_8
000	0	0	0	0	0	0	0	0
100	1	1	1	1	1	0	1	1
010	0	01	01	01	01	0	0	0
110	1	11	11	11	11	0	1	1
001	01	0	01	1	11	1	0	1
101	11	1	11	0	01	1	1	0
011	01	01	0	11	1	1	0	1
111	11	11	1	01	0	1	1	0

Table: A $(8,|C|,4)$ code constructed from $\mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_2$, $|C| = 8$. Pairs are elements in $\mathbb{Z}_2 \oplus \mathbb{Z}_2$.

- It is possible to construct a $(n, |C|, d) = (2^{k+1} + k - 2, 2^{k+1}, 2^k)$ quasi-uniform code from $\mathbb{Z}_2 \oplus \ldots \oplus \mathbb{Z}_2$ of order 2^{k+1}.

Some Properties

Let G be a finite group with subgroups $G_1, \ldots, G_n$.

Lemma (Size)

The size of the quasi-uniform code $|C| = |G|/|G_{\mathcal{N}}|$; $G_{\mathcal{N}} = \cap_{i=1}^{n} G_i$.

Some Properties

Let G be a finite group with subgroups $G_1, \ldots, G_n$.

Lemma (Size)

The size of the quasi-uniform code $|C| = |G|/|G_{\mathcal{N}}|$; $G_{\mathcal{N}} = \cap_{i=1}^{n} G_i$.

Proposition (Group Structure)

Quasi-uniform code C has a group structure if all subgroups G_i are normal.

Some Properties

Let G be a finite group with subgroups $G_1, \ldots, G_n$.

Lemma (Size)

The size of the quasi-uniform code $|C| = |G|/|G_{\mathcal{N}}|$; $G_{\mathcal{N}} = \cap_{i=1}^{n} G_i$.

Proposition (Group Structure)

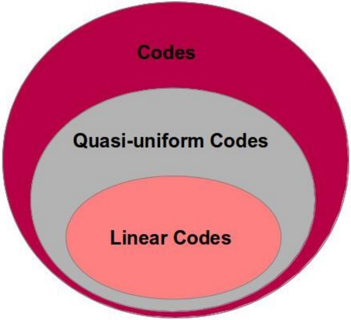
Quasi-uniform code C *has a group structure if all subgroups* G_i *are normal.*

Lemma (Minimum Distance)

The minimum distance of the quasi-uniform code (having group structure) C *is*

$$n - \max_{\mathcal{A} \in \mathcal{N}, G_{\mathcal{A}} \neq \{0\}} |\mathcal{A}|.$$

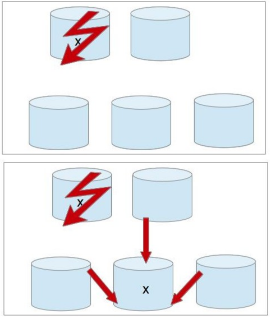
Quasi-uniform codes:Overview



Using this method, one can construct:

- Codes over non-field alphabets.
- Codes with coefficients over different alphabets.
- Non-linear codes.

Storage code construction



- Encode a data object into a codeword, spread the coefficients across nodes.
- The code provides fault tolerance in case of node failures.
- Needs to enable repairability.

Suppose a data object (u_1, u_2, u_3) needs to be stored. Consider the following storage allocation:

node 1: (u_1, u_3)	node 5: $(u_1 + u_3, u_2 + u_3)$
node 2: (u_1, u_2)	node 6: $(u_3, 0)$
node 3: $(u_1, u_2 + u_3)$	node 7: $(u_1, 0)$
node 4: $(u_1 + u_3, u_2)$	node 8: $(u_1 + u_3, 0)$

Three failures at most can be tolerated (corresponding to a minimum distance of 4 indeed).

In case of one node failure, this node is repaired easily: indeed, this codeword is created from $\mathbb{Z}_2$-linear combinations. For example:

- node 1 is repaired by downloading u_1 (from node 2 or node 7) and u_3 (from node 6),
- node 2 is repaired by downloading u_2 from node 4 and u_1 (from node 1 or 7).

Codes with locality and availability

- A code of length n and dimension k is said to be (n, k, r) locally repairable, if each codeword symbol can be recovered from r other symbols. The integer $r; 1 \leq r \leq k$, is called locality.
- If there exist t disjoint recovery sets for each codeword symbol, the code is said to have availability t.
- The above construction gives codes with locality and availability, where $t = r = 2$.
- It satisfies some bounds for codes with locality and availability.

For more detail, please refer:
`http://ieeexplore.ieee.org/document/6620274/`
`http://ieeexplore.ieee.org/abstract/document/6983940/`

Thank You!!