Coding problems for memory and storage applications

Alexander Barg

University of Maryland

January 27, 2015

Introduction: Big Data

Big Data players: Facebook, Instagram, Google, MSFT, etc.; Dropbox, Box, etc.
Companies marketing coding solutions: CleverSafe (RS codes) and others.

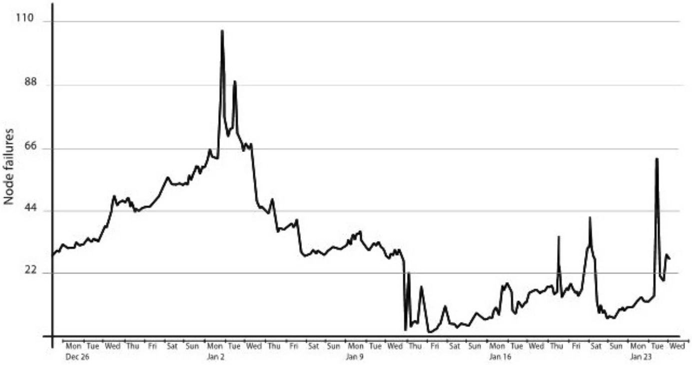
Introduction: Big Data

Big Data players: Facebook, Instagram, Google, MSFT, etc.; Dropbox, Box, etc.
Companies marketing coding solutions: CleverSafe (RS codes) and others.

Cluster of machines running Hadoop at Yahoo!

Node failures are the **norm**

Is repair cost a real issue?

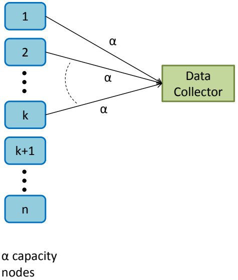


(Average number of failed nodes =20) $\times$15Tb $=$ 300Tb

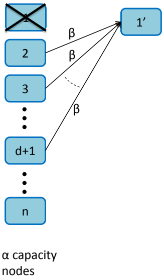
Two approaches to data coding in distributed storage

- Codes with locality
- Regenerating codes

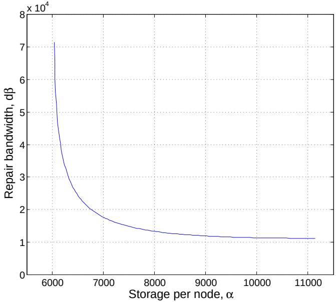
Regenerating codes



Data collection



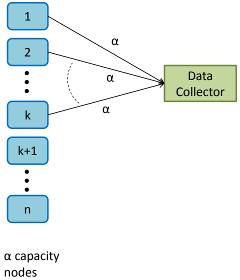
Node repair



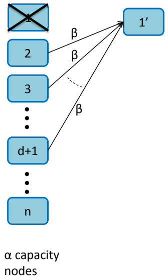
Trade-off

- B symbols are encoded into $n\alpha$ symbols stored in n nodes

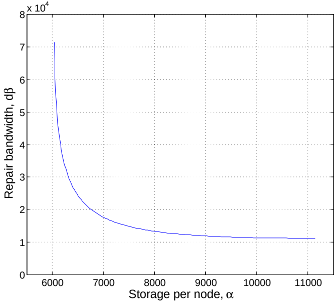
Regenerating codes



Data collection



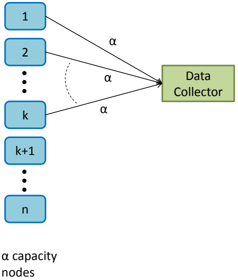
Node repair



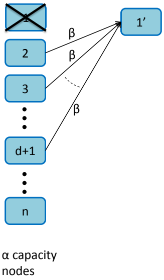
Trade-off

- B symbols are encoded into $n\alpha$ symbols stored in n nodes
- Downloading the data is possible by accessing any k nodes

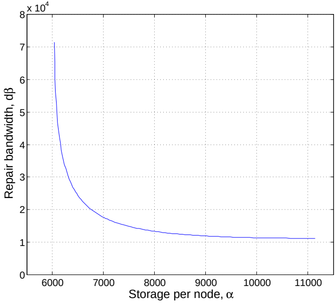
Regenerating codes



Data collection



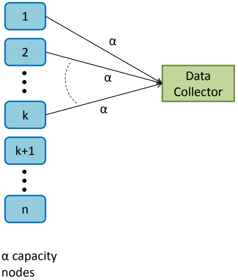
Node repair



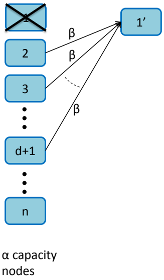
Trade-off

- *B* symbols are encoded into $n\alpha$ symbols stored in *n* nodes
- Downloading the data is possible by accessing any *k* nodes
- Node repair (exact or functional) can be performed by downloading $\beta < \alpha$ symbols from any subset of *d* nodes.

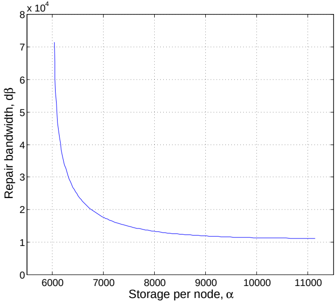
Regenerating codes



Data collection



Node repair

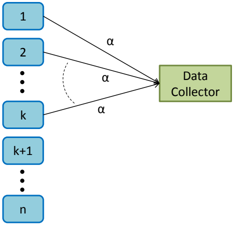


Trade-off

- *B* symbols are encoded into $n\alpha$ symbols stored in *n* nodes
- Downloading the data is possible by accessing any *k* nodes
- Node repair (exact or functional) can be performed by downloading $\beta < \alpha$ symbols from any subset of *d* nodes.
- Repair bandwidth $d\beta$

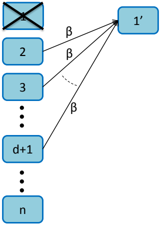
$(n, k, d, \{\alpha, \beta\})$ regenerating codes

Regenerating codes



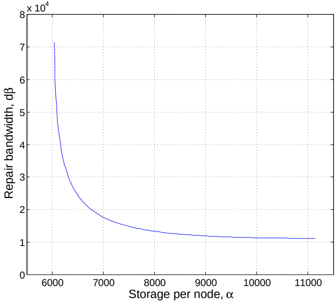
α capacity nodes

Data collection



α capacity nodes

Node repair



Trade-off

- B symbols are encoded into $n\alpha$ symbols stored in n nodes
- Downloading the data is possible by accessing any k nodes
- Node repair (exact or functional) can be performed by downloading $\beta < \alpha$ symbols from any subset of d nodes.
- Repair bandwidth $d\beta$

$(n, k, d, \{\alpha, \beta\})$ regenerating codes

A. Dimakis, P. Godfrey, Y. Wu, M. Wainwright, and **K. Ramchandran**, Network coding for distributed storage systems, 2010

Locally recoverable codes: Plan

In this part we focus on locally recoverable codes

Locally recoverable codes: Plan

In this part we focus on locally recoverable codes

LRC code: To recover one lost symbol of the encoding it suffices to access a small number r of other symbols.

Locally recoverable codes: Plan

In this part we focus on locally recoverable codes

LRC code: To recover one lost symbol of the encoding it suffices to access a small number r of other symbols.

1. Current solutions
2. Parameters of LRC codes
3. MDS-like codes with the locality property
4. The availability problem: Multiple recovering sets
5. Extensions
   - LRC codes on algebraic curves
   - Cyclic LRC codes
6. Open problems: Bounds on codes; cyclic codes; list decoding

State-of-the-Art Coding technique

RAID: Redundant Array of Independent Disks

RAID 1 – Replication (currently 3x)

- Provides high availability of information
- Can tolerate any 2 disk failures
- Widely used in *Hadoop* and many other systems
- Storage overhead of 200%

RAID 6 uses [6,4,3] RS codes

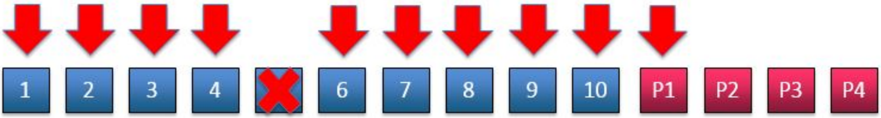
$[n, k]$ RS codes

- Can tolerate any n-k disk failures
- Poor handling of single disk failures (The Repair Problem)

Limitations of Reed-Solomon codes

Example: $[14, 10]$ RS code

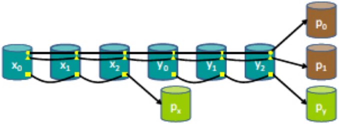
Transmit 10 symbols to recover one lost value



Generates 10x more traffic for recovery of one drive
If large portion of the cluster is RS-coded, this leads to saturation of the network

Other constructions

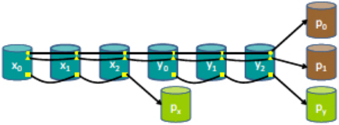
A combination of local and global parity checks for single and multiple nodes failures



(**C. Huang** at al., Erasure coding in Windows Azure Storage, USENIX Conf. 2012)

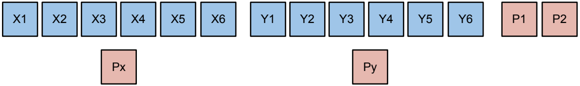
Other constructions

A combination of local and global parity checks for single and multiple nodes failures



(**C. Huang** at al., Erasure coding in Windows Azure Storage, USENIX Conf. 2012)

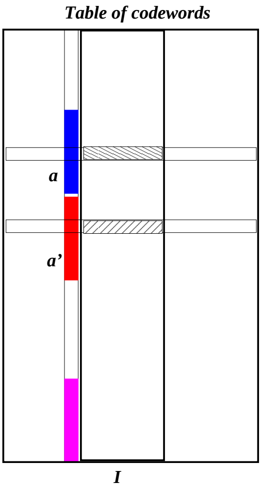
Other similar constructions (Windows Azure code)



Pyramid codes (**C. Huang** et al., 2007)

Locally recoverable codes

The code $\mathcal{C} \subset \mathbb{F}^n$ is locally recoverable with locality r if every symbol can be recovered by accessing some other r symbols in the encoding (recovering set of coordinate i)



(n, k, r) LRC code

Let $a \in \mathbb{F}$; consider the restriction $\mathcal{C}_J$ of $\mathcal{C}$ to a subset $J \subset [n]$.
Let

$$\mathcal{C}_J(a, i) = \{x \in \mathcal{C}_J : x_i = a\}, \quad i \in [n].$$

Definition

Code $\mathcal{C}$ has *locality* r if for every $i \in [n]$ there exists a subset $J_i \subset [n] \backslash i, |J_i| \leq r$ such that

$$\mathcal{C}_{J_i}(a, i) \cap \mathcal{C}_{J_i}(a', i) = \emptyset, \quad a \neq a'$$

(n, k, r) LRC code

Let $a \in \mathbb{F}$; consider the restriction $\mathcal{C}_J$ of $\mathcal{C}$ to a subset $J \subset [n]$.
Let

$$\mathcal{C}_J(a, i) = \{x \in \mathcal{C}_J : x_i = a\}, \quad i \in [n].$$

Definition

Code $\mathcal{C}$ has *locality r* if for every $i \in [n]$ there exists a subset $J_i \subset [n] \backslash i, |J_i| \leq r$ such that

$$\mathcal{C}_{J_i}(a, i) \cap \mathcal{C}_{J_i}(a', i) = \emptyset, \quad a \neq a'$$

J. Han and **L. Lastras-Montano**, *ISIT* 2007;
C. Huang, M. Chen, and **J. Li**, *Symp. Networks App.* 2007;
F. Oggier and **A. Datta** '10;
P. Gopalan, C. Huang, H. Simitci, and **S. Yekhanin**, *IEEE Trans. Inf. Theory,* Nov. 2012.

Parameters of LRC codes

Parameters of LRC codes

Theorem

Let $\mathcal{C}$ be an (n, k, r) LRC code of cardinality q^k over an alphabet of size q, then:
The rate of $\mathcal{C}$ satisfies

$$\frac{k}{n} \le \frac{r}{r+1}. \tag{1}$$

The minimum distance of $\mathcal{C}$ satisfies

$$d \le n - k - \left\lceil \frac{k}{r} \right\rceil + 2. \tag{2}$$

Bound (2) is due to **Gopalan** e.a. (2011) and **Papailiopoulos** e.a. (2012).

Parameters of LRC codes

Theorem

Let $\mathcal{C}$ be an (n, k, r) LRC code of cardinality q^k over an alphabet of size q, then:
The rate of $\mathcal{C}$ satisfies

$$\frac{k}{n} \le \frac{r}{r+1}. \tag{1}$$

The minimum distance of $\mathcal{C}$ satisfies

$$d \le n - k - \left\lceil \frac{k}{r} \right\rceil + 2. \tag{2}$$

Bound (2) is due to **Gopalan** e.a. (2011) and **Papailiopoulos** e.a. (2012).

Note that $r = k$ reduces (2) to the Singleton bound

$$d \le n - k + 1$$

The distance bound

Main idea. Let $\mathcal{C}$ be a q-ary code of length n, size q^k. The distance $d(\mathcal{C})$ equals

$$d(\mathcal{C}) = n - \max_{S\subset[n]}\{|S| : |\mathcal{C}_S| < q^k\}$$

The distance bound

Main idea. Let $\mathcal{C}$ be a q-ary code of length n, size q^k. The distance $d(\mathcal{C})$ equals

$$d(\mathcal{C}) = n - \max_{S \subset [n]} \{|S| : |\mathcal{C}_S| < q^k\}$$

The Singleton bound (without locality): $|S| = k - 1$

The distance bound

Main idea. Let $\mathcal{C}$ be a q-ary code of length n, size q^k. The distance $d(\mathcal{C})$ equals

$$d(\mathcal{C}) = n - \max_{S\subset[n]}\{|S| : |\mathcal{C}_S| < q^k\}$$

The Singleton bound (with locality):

Let $I_i \subset [n], |I_i| \le r$ be the recovering set for the symbol c_i, $i = 1, ..., n$.

The distance bound

Main idea. Let $\mathcal{C}$ be a q-ary code of length n, size q^k. The distance $d(\mathcal{C})$ equals

$$d(\mathcal{C}) = n - \max_{S\subset[n]}\{|S| : |\mathcal{C}_S| < q^k\}$$

The Singleton bound (with locality):

Let $I_i \subset [n], |I_i| \le r$ be the recovering set for the symbol $c_i, i = 1, ..., n$.

Let $J_m = \cup_{i=1}^m I_i$, where $m = \lfloor (k-1)/r \rfloor$. Clearly $|J_m| \le k - 1$.

The distance bound

Main idea. Let $\mathcal{C}$ be a q-ary code of length n, size q^k. The distance $d(\mathcal{C})$ equals

$$d(\mathcal{C}) = n - \max_{S\subset[n]}\{|S| : |\mathcal{C}_S| < q^k\}$$

The Singleton bound (with locality):

Let $I_i \subset [n], |I_i| \le r$ be the recovering set for the symbol $c_i, i = 1, ..., n$.

Let $J_m = \cup_{i=1}^m I_i$, where $m = \lfloor (k-1)/r \rfloor$. Clearly $|J_m| \le k-1$.

Consider the subset $J_m' = J_m \cup \{1, \dots, m\}$. We have $\mathcal{C}_{J_m'} \le q^{k-1}$.

The distance bound

Main idea. Let $\mathcal{C}$ be a q-ary code of length n, size q^k. The distance $d(\mathcal{C})$ equals

$$d(\mathcal{C}) = n - \max_{S \subset [n]} \{|S| : |\mathcal{C}_S| < q^k\}$$

The Singleton bound (with locality):

Let $I_i \subset [n], |I_i| \le r$ be the recovering set for the symbol $c_i, i = 1, ..., n$.

Let $J_m = \cup_{i=1}^m I_i$, where $m = \lfloor (k-1)/r \rfloor$. Clearly $|J_m| \le k - 1$.

Consider the subset $J_m' = J_m \cup \{1, \dots, m\}$. We have $\mathcal{C}_{J_m'} \le q^{k-1}$.

If $|J_m'| < k - 1$, add to J_m' any $k - 1 - |J_m|$ other coordinates to form the set $L_m \subset [n]$.

The distance bound

Main idea. Let $\mathcal{C}$ be a q-ary code of length n, size q^k. The distance $d(\mathcal{C})$ equals

$$d(\mathcal{C}) = n - \max_{S\subset[n]}\{|S| : |\mathcal{C}_S| < q^k\}$$

The Singleton bound (with locality):

Let $I_i \subset [n], |I_i| \le r$ be the recovering set for the symbol $c_i, i = 1, ..., n$.

Let $J_m = \cup_{i=1}^m I_i$, where $m = \lfloor (k-1)/r \rfloor$. Clearly $|J_m| \le k-1$.

Consider the subset $J'_m = J_m \cup \{1, \dots, m\}$. We have $\mathcal{C}_{J'_m} \le q^{k-1}$.

If $|J'_m| < k-1$, add to J'_m any $k-1-|J_m|$ other coordinates to form the set $L_m \subset [n]$.

We have

$$|\mathcal{C}_{L_m}| < q^k$$

$$|L_m| = k-1+m = k-1+\left\lfloor \frac{k-1}{r} \right\rfloor = k-2+\left\lceil \frac{k}{r} \right\rceil$$

Cadambe-Mazumdar bound

(n, k, r) LRC code $\mathcal{C}$

Cadambe-Mazumdar bound

(n, k, r) LRC code $\mathcal{C}$

$$\boxed{k \le \min_{s \ge 1}(rs + k_{\text{opt}}^{(q)}(n - s(r+1), d))}$$

Cadambe-Mazumdar bound

(n, k, r) LRC code $\mathcal{C}$

$$\boxed{k \le \min_{s \ge 1}(rs + k_{\text{opt}}^{(q)}(n - s(r+1), d))}$$

Consider the sets of coordinates L_s constructed above, $1 \le s \le \lfloor (k-1)/r \rfloor$.

$$|\mathcal{C}_{L_s}| \le q^{rs}$$

The shortening of the code $\mathcal{C}$ on the coordinates in L_s forms a code of length $n - s(r+1)$ with distance d

Existence (Gilbert-Varshamov) bound

A linear q-ary $[n, k', d]$ code exists if

$$\sum_{i=0}^{d-2} \binom{n-1}{i} (q-1)^i < q^{n-k'}$$

Add $\lceil n/(r+1) \rceil$ local parities

$$k \geq k' - \left\lceil \frac{n}{r+1} \right\rceil$$

Sequences of (R, δ) codes with locality r exist as long as

$$R < \frac{r}{r+1} - \delta \log_q \frac{q-1}{\delta} - (1-\delta) \log_q \frac{1}{1-\delta}$$

Existence (Gilbert-Varshamov) bound

A linear q-ary $[n, k', d]$ code exists if

$$\sum_{i=0}^{d-2} \binom{n-1}{i} (q-1)^i < q^{n-k'}$$

Add $\lceil n/(r+1) \rceil$ local parities

$$k \geq k' - \left\lceil \frac{n}{r+1} \right\rceil$$

Sequences of (R, δ) codes with locality r exist as long as

$$\boxed{R < \frac{r}{r+1} - \delta \log_q \frac{q-1}{\delta} - (1-\delta) \log_q \frac{1}{1-\delta}}$$

$$R \leq \frac{r}{r+1} - h_q(\delta)$$

Early constructions

1. Optimal $((r+1)\lceil k/r\rceil, k, r)$ LRC code

   Prasanth, Kamath, Lalitha, and **Kumar**, ISIT 2012
   Restricted length

2. Optimal (n, k, r) LRC codes

   Silberstein, Rawat, Koluoglu, and **Vishwanath**, ISIT 2013
   Tamo, Papailiopoulos, and **Dimakis**, ISIT 2013

   Almost any n, k, r

   Field size $q \sim 2^n$

Reed-Solomon codes

An RS code $\mathcal{C}$ is a linear code of length $n \le q-1$ over the field $\mathbb{F}_q$

Reed-Solomon codes

An RS code $\mathcal{C}$ is a linear code of length $n \leq q-1$ over the field $\mathbb{F}_q$

Given a polynomial $f \in \mathbb{F}_q[x]$ and a set $A = \{P_1, \dots, P_n\} \subset \mathbb{F}_q$ define the map

$$ev_A : f \mapsto (f(P_i), i = 1, \dots, n)$$

Reed-Solomon codes

An RS code $\mathcal{C}$ is a linear code of length $n \le q-1$ over the field $\mathbb{F}_q$

Given a polynomial $f \in \mathbb{F}_q[x]$ and a set $A = \{P_1, \dots, P_n\} \subset \mathbb{F}_q$ define the map

$$ev_A : f \mapsto (f(P_i), i = 1, \dots, n)$$

RS code $\mathcal{C}$ encodes messages of k symbols.
Let $V_k(q) = \{f \in \mathbb{F}_q[x] : \deg(f) \le k-1\}$

$$\mathcal{C} : V_k(q) \to \mathbb{F}_q^n$$
$$f \mapsto ev_A(f) = (f(P_i), i = 1, \dots, n)$$

Reed-Solomon codes

An RS code $\mathcal{C}$ is a linear code of length $n \le q-1$ over the field $\mathbb{F}_q$

Given a polynomial $f \in \mathbb{F}_q[x]$ and a set $A = \{P_1, \dots, P_n\} \subset \mathbb{F}_q$ define the map

$$ev_A : f \mapsto (f(P_i), i = 1, \dots, n)$$

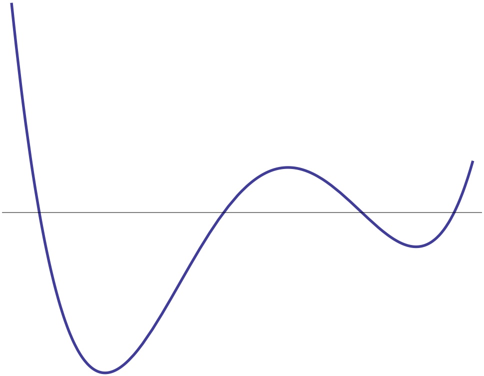
RS code $\mathcal{C}$ encodes messages of k symbols.
Let $V_k(q) = \{f \in \mathbb{F}_q[x] : \deg(f) \le k-1\}$

$$\begin{aligned} \mathcal{C} : V_k(q) &\to \mathbb{F}_q^n \\ f &\mapsto ev_A(f) = (f(P_i), i = 1, \dots, n) \end{aligned}$$

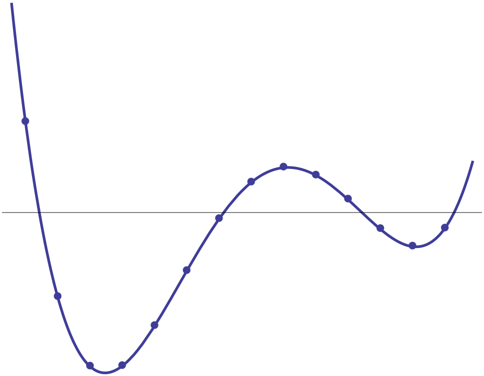
Example: Let $q = 8$, $f(x) = 1 + \alpha x + \alpha x^2$

$$f(x) \mapsto (1, \alpha^4, \alpha^6, \alpha^4, \alpha, \alpha, \alpha^6)$$

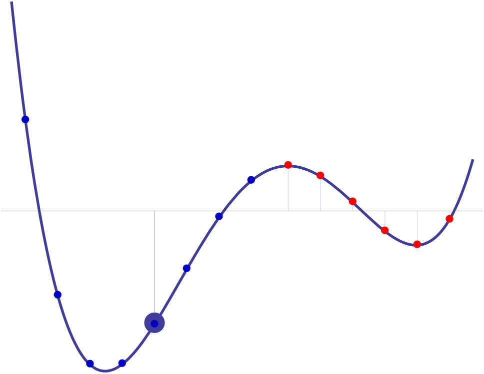
Reed-Solomon codes



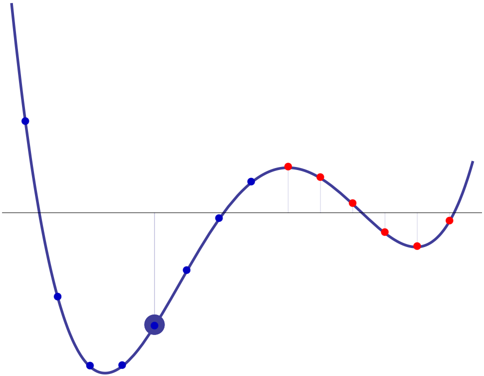
Reed-Solomon codes



Reed-Solomon codes



Reed-Solomon codes



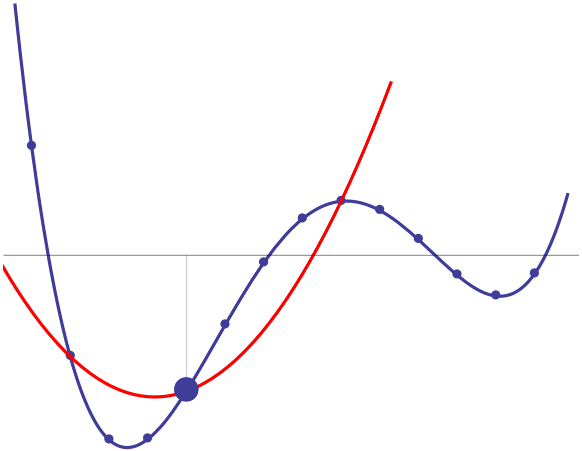
To recover one erased value we need to read *k* other values

LRC codes: Idea of construction

What if we can interpolate low-degree polynomials?

LRC codes: Idea of construction

What if we can interpolate low-degree polynomials?



Construction of LRC codes: Limitations

We need a specially chosen set of points A

Restricted set of polynomials

Construction of (n, k, r) LRC codes: Example

Parameters: $n = 9, k = 4, r = 2, q = 13$;

Set of points: A={1,2,3,4,5,6,9,10,12}

$$\mathcal{A} = \{A_1 = \{1, 3, 9\}, A_2 = \{2, 6, 5\}, A_3 = \{4, 12, 10\}\}$$

Message: $a = (a_{0,0}, a_{0,1}, a_{1,0}, a_{1,1}) \in \mathbb{F}_q^k$

Polynomial space:

$$V_k(q) := \{a_{0,0} + a_{1,0}x + a_{0,1}x^3 + a_{1,1}x^4\}$$

Construction of (n, k, r) LRC codes: Example

Parameters: $n = 9, k = 4, r = 2, q = 13$;

Set of points: A={1,2,3,4,5,6,9,10,12}

$$\mathcal{A} = \{A_1 = \{1, 3, 9\}, A_2 = \{2, 6, 5\}, A_3 = \{4, 12, 10\}\}$$

Message: $a = (a_{0,0}, a_{0,1}, a_{1,0}, a_{1,1}) \in \mathbb{F}_q^k$

Polynomial space:

$$V_k(q) := \{a_{0,0} + a_{1,0}x + a_{0,1}x^3 + a_{1,1}x^4\}$$

E.g., $a = (1, 1, 1, 1)$, $f_a(x) = 1 + x + x^3 + x^4$; $ev_A(f) = (4, 8, 7, 1, 11, 2, 0, 0, 0)$

Construction of (n, k, r) LRC codes: Example

Parameters: $n = 9, k = 4, r = 2, q = 13;$

Set of points: A={1,2,3,4,5,6,9,10,12}

$$\mathcal{A} = \{A_1 = \{1, 3, 9\}, A_2 = \{2, 6, 5\}, A_3 = \{4, 12, 10\}\}$$

Message: $a = (a_{0,0}, a_{0,1}, a_{1,0}, a_{1,1}) \in \mathbb{F}_q^k$

Polynomial space:

$$V_k(q) := \{a_{0,0} + a_{1,0}x + a_{0,1}x^3 + a_{1,1}x^4\}$$

E.g., $a = (1, 1, 1, 1)$, $f_a(x) = 1 + x + x^3 + x^4$; $ev_A(f) = (4, 8, 7, 1, 11, 2, 0, 0, 0)$

Say $c_1 = f_a(1)$ is erased. We access the recovering set A_1 to construct a line $\delta(x) = 2x + 2$ such that $\delta(3) = 8, \delta(9) = 7$.

Construction of (n, k, r) LRC codes: Example

Parameters: $n = 9, k = 4, r = 2, q = 13$;

Set of points: A={1,2,3,4,5,6,9,10,12}

$$\mathcal{A} = \{A_1 = \{1, 3, 9\}, A_2 = \{2, 6, 5\}, A_3 = \{4, 12, 10\}\}$$

Message: $a = (a_{0,0}, a_{0,1}, a_{1,0}, a_{1,1}) \in \mathbb{F}_q^k$

Polynomial space:

$$V_k(q) := \{a_{0,0} + a_{1,0}x + a_{0,1}x^3 + a_{1,1}x^4\}$$

E.g., $a = (1, 1, 1, 1)$, $f_a(x) = 1 + x + x^3 + x^4$; $ev_A(f) = (4, 8, 7, 1, 11, 2, 0, 0, 0)$

Say $c_1 = f_a(1)$ is erased. We access the recovering set A_1 to construct a line $\delta(x) = 2x + 2$ such that $\delta(3) = 8, \delta(9) = 7$.

Compute c_1 as $\delta(1) = 4$

Construction of (n, k, r) LRC codes: Example

Parameters: $n = 9, k = 4, r = 2, q = 13$;

Set of points: A={1,2,3,4,5,6,9,10,12}

$$\mathcal{A} = \{A_1 = \{1, 3, 9\}, A_2 = \{2, 6, 5\}, A_3 = \{4, 12, 10\}\}$$

Message: $a = (a_{0,0}, a_{0,1}, a_{1,0}, a_{1,1}) \in \mathbb{F}_q^k$

Polynomial space:

$$V_k(q) := \{a_{0,0} + a_{1,0}x + a_{0,1}x^3 + a_{1,1}x^4\}$$

E.g., $a = (1, 1, 1, 1)$, $f_a(x) = 1 + x + x^3 + x^4$; $ev_A(f) = (4, 8, 7, 1, 11, 2, 0, 0, 0)$

Say $c_1 = f_a(1)$ is erased. We access the recovering set A_1 to construct a line $\delta(x) = 2x + 2$ such that $\delta(3) = 8, \delta(9) = 7$.

Compute c_1 as $\delta(1) = 4$

It works!

Construction of (n, k, r) LRC codes

Assume that $q \geq n, (r+1)|n, r|k$
Let $A \subseteq \mathbb{F}_q, |A| = n$

Construction of (n, k, r) LRC codes

Assume that $q \geq n, (r+1)|n, r|k$
Let $A \subseteq \mathbb{F}_q, |A| = n$

Suppose there exists a polynomial $g(x) \in \mathbb{F}[x]$ such that

1. $\deg g = r + 1$,
2. There exists a partition $\mathcal{A} = \{A_1, ..., A_{\frac{n}{r+1}}\}$ of A into sets of size $r+1$, such that g is constant on each set A_i in the partition. For all $i = 1, \dots, n/(r+1)$, and any $\alpha, \beta \in A_i$,
$$g(\alpha) = g(\beta).$$

Construction of (n, k, r) LRC codes

Assume that $q \geq n, (r+1)|n, r|k$
Let $A \subseteq \mathbb{F}_q, |A| = n$

Suppose there exists a polynomial $g(x) \in \mathbb{F}[x]$ such that

1. $\deg g = r + 1$,
2. There exists a partition $\mathcal{A} = \{A_1, ..., A_{\frac{n}{r+1}}\}$ of A into sets of size $r+1$, such that g is constant on each set A_i in the partition. For all $i = 1, \dots, n/(r+1)$, and any $\alpha, \beta \in A_i$,
$$g(\alpha) = g(\beta).$$

E.g., $n = 9, r = 2, q = 13$;
$$\mathcal{A} = \{A_1 = \{1, 3, 9\}, A_2 = \{2, 6, 5\}, A_3 = \{4, 12, 10\}\},$$
Then $g(x) = x^3$ is constant on each of the A_i's

Construction of (n, k, r) LRC codes

Given $A \subset \mathbb{F}$, partition $\mathcal{A}$ into $(r+1)$-subsets.

To encode the message $a \in \mathbb{F}^k$, write $a = (a_{ij}, i = 0, \ldots, r-1; j = 0, ..., \frac{k}{r} - 1)$

Define the encoding polynomial

$$f_a(x) = \sum_{i=0}^{r-1} f_i(x) x^i,$$

where

$$f_i(x) = \sum_{j=0}^{\frac{k}{r}-1} a_{ij} g(x)^j, \quad i = 0, ..., r-1$$

A linear code $\mathcal{C}$ is constructed as follows:

$$\begin{aligned} Ev : & \mathbb{F}^k \to \mathbb{F}^n \\ & a \mapsto (f_a(\beta), \beta \in A) \end{aligned}$$

Recovery of erased symbol

Suppose that the location of erased symbol is $\alpha \in A_j$; $A_j \in \mathcal{A}$

To find c_α we rely on the recovering set A_j

Find a polynomial $\delta(x)$ s.t. $\delta(\beta) = c_\beta, \beta \in A_j \backslash \alpha$; $\deg \delta \leq r - 1$:

$$\delta(x) = \sum_{\beta \in A_j \backslash \alpha} c_\beta \prod_{\beta' \in A_j \backslash \{\alpha, \beta\}} \frac{x - \beta'}{\beta - \beta'}$$

Then $c_\alpha = \delta(\alpha)$

Properties of the construction

Theorem

The constructed linear codes are optimal (n, k, r) LRC codes with respect to the "Singleton bound" (2).

Optimality is proved by counting degrees.

Locality: Let $\alpha \in A_j$ be the erased location. Define

$$\partial(x) = \sum_{i=0}^{r-1} f_i(\alpha)x^i$$

By the construction, for all $\beta \in A_j$

$$\partial(\beta) = f_a(\beta)$$

Since $\deg \partial \leq r - 1$, we see that $\partial(x) \equiv \delta(x)$.

Constructing the polynomial $g(x)$

Proposition

Let H be a subgroup of $\mathbb{F}_q^*$ or $\mathbb{F}_q^+$. The annihilator polynomial of H

$$g(x) = \prod_{h \in H} (x - h)$$

is constant on each coset of H.

Constructing the polynomial $g(x)$

Proposition

Let H be a subgroup of $\mathbb{F}_q^*$ or $\mathbb{F}_q^+$. The annihilator polynomial of H

$$g(x) = \prod_{h \in H} (x - h)$$

is constant on each coset of H.

Assume that H is a multiplicative subgroup and let $a, a\bar{h}$ be two elements of the coset aH, where $\bar{h} \in H$, then

$$\begin{aligned} g(a\bar{h}) = \prod_{h \in H} (a\bar{h} - h) &= \bar{h}^{|H|} \prod_{h \in H} (a - h\bar{h}^{-1}) \\ &= \prod_{h \in H} (a - h) \\ &= g(a). \end{aligned}$$

Some generalizations

The locator set $A \subset \mathbb{F}$, $A = \sqcup_{i=1}^{m} A_i$. Consider the algebra

$$\mathbb{F}_{\mathcal{A}}[x] = \{f \in \mathbb{F}[x] : f \text{ is constant on } A_i, i = 1, \dots, m;\ \deg f < |A|\}.$$

Some generalizations

The locator set $A \subset \mathbb{F}$, $A = \sqcup_{i=1}^{m} A_i$. Consider the algebra

$$\mathbb{F}_{\mathcal{A}}[x] = \{f \in \mathbb{F}[x] : f \text{ is constant on } A_i, i = 1, \dots, m;\ \deg f < |A|\}.$$

The properties of $\mathbb{F}_{\mathcal{A}}[x]$ are summarized as follows:

1. $\dim(\mathbb{F}_{\mathcal{A}}[x]) = m$;
2. Let $\alpha_1, ..., \alpha_m$ be distinct nonzero elements of $\mathbb{F}$, and let g be the polynomial of degree $\deg(g) < |A|$ that satisfies $g(A_i) = \alpha_i$ for all $i = 1, ..., m$. Then the polynomials $1, g, ..., g^{m-1}$ form a basis of $\mathbb{F}_{\mathcal{A}}[x]$.

Some generalizations

The locator set $A \subset \mathbb{F}$, $A = \sqcup_{i=1}^{m} A_i$. Consider the algebra

$$\mathbb{F}_{\mathcal{A}}[x] = \{f \in \mathbb{F}[x] : f \text{ is constant on } A_i, i = 1, \dots, m; \ \deg f < |A|\}.$$

The properties of $\mathbb{F}_{\mathcal{A}}[x]$ are summarized as follows:

1. $\dim(\mathbb{F}_{\mathcal{A}}[x]) = m$;
2. Let $\alpha_1, ..., \alpha_m$ be distinct nonzero elements of $\mathbb{F}$, and let g be the polynomial of degree $\deg(g) < |A|$ that satisfies $g(A_i) = \alpha_i$ for all $i = 1, ..., m$. Then the polynomials $1, g, ..., g^{m-1}$ form a basis of $\mathbb{F}_{\mathcal{A}}[x]$.

General code construction: Let $A \subset \mathbb{F}$, $|A| = n$; $A = \sqcup_{i=1}^{m} A_i$, $|A_i| = r + 1$ for all i. Let Φ be an injective mapping from $\mathbb{F}^k$ to the space of polynomials

$$\mathcal{F}_{\mathcal{A}}^{r} = \oplus_{i=0}^{r-1} \mathbb{F}_{\mathcal{A}}[x] x^i.$$

The evaluation code obtained in this way is an (n, k, r) LRC code.

Extensions

1. It is possible to lift the divisibility constraints $r|k, (r+1)|n$

Extensions

1. It is possible to lift the divisibility constraints $r|k, (r+1)|n$
2. It is possible to define a systematic algebraic encoding mapping.

Extensions

1. It is possible to lift the divisibility constraints $r|k, (r+1)|n$
2. It is possible to define a systematic algebraic encoding mapping.
3. To improve data availability, replace $[r+1, r, 2]$ local codes with $[r+\rho-1, r]$ MDS codes. Then every c_i is a function of any r out of $r+\rho-1$ coordinates.
   Bound on the distance:

$$d \le n - k + 1 - \left(\left\lceil \frac{k}{r} \right\rceil - 1\right)(\rho - 1) \quad \text{(Kamath e.a., 2013)}$$

Extensions

1. It is possible to lift the divisibility constraints $r|k, (r+1)|n$
2. It is possible to define a systematic algebraic encoding mapping.
3. To improve data availability, replace $[r+1, r, 2]$ local codes with $[r+\rho-1, r]$ MDS codes. Then every c_i is a function of any r out of $r+\rho-1$ coordinates.
   Bound on the distance:

$$d \le n - k + 1 - \left(\left\lceil \frac{k}{r} \right\rceil - 1\right)(\rho - 1) \quad \text{(Kamath e.a., 2013)}$$

Claim: Taking recovering sets of size $|A_i| = r + \rho - 1$ and a polynomial basis of $\mathbb{F}_{\mathcal{A}}[x]$, we can construct an (n, k, r) LRC code whose distance meets this bound.

Availability problem

"Hot data" accessed simultaneously by a very large number of users

Availability problem

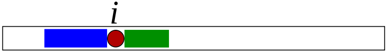
"Hot data" accessed simultaneously by a very large number of users

Multiple recovering sets: Definition

Every symbol in data encoding appears in several disjoint (orthogonal) parity checks

$\mathcal{C} \subset \mathbb{F}^n$ a code of length n

Every coordinate is recoverable from the codeword symbols in several recovering sets:



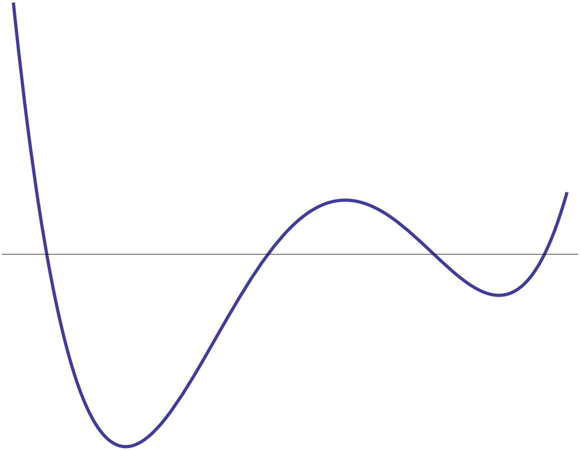
Multiple recovering sets: Definition

Let $\mathcal{C}(a,i) = \{x \in \mathcal{C} : x_i = a\}, a \in \mathbb{F}, i \in [n]$

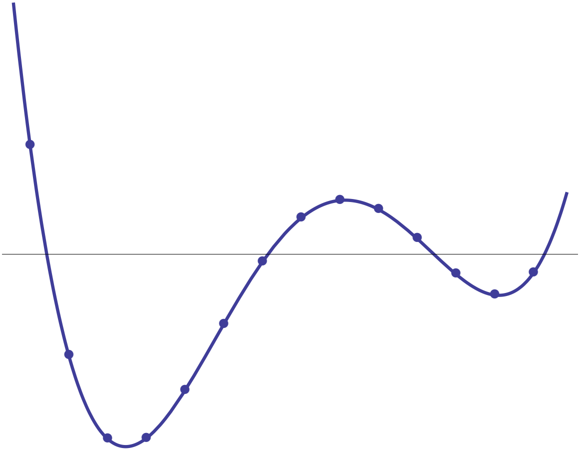
The code $\mathcal{C}$ has two disjoint recovering sets if for every $i \in [n]$ there are subsets $R_i^1, R_i^2 \subset [n]\backslash\{i\}, R_i^1 \cap R_i^2 = \emptyset$ such that

$$\mathcal{C}(a,i)_{R_i^j} \cap \mathcal{C}(a',i)_{R_i^j} = \emptyset, \quad a \neq a';\ j = 1,2$$

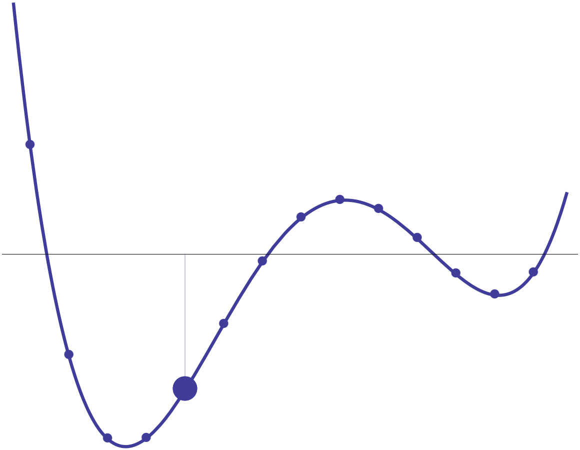
Multiple recovering sets: Idea of construction



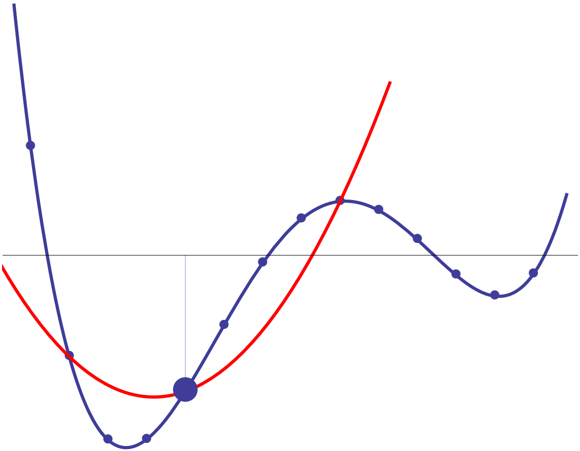
Multiple recovering sets: Idea of construction



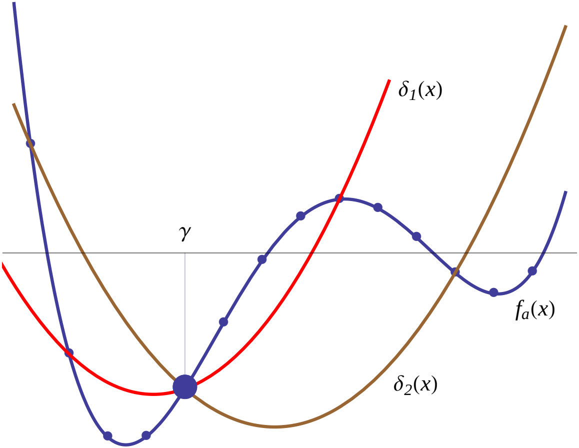
Multiple recovering sets: Idea of construction



Multiple recovering sets: Idea of construction



Multiple recovering sets: Idea of construction



$f_a(\gamma)$ can be found by interpolating $\delta_1(x)$ as well as $\delta_2(x)$

Multiple recovering sets: Example

Take $\mathbb{F} = \mathbb{F}_{13}$; $G, H \le \mathbb{F}^*$; $G = \langle 5 \rangle, H = \langle 3 \rangle$

$$\mathcal{A}_G = \{\{1, 5, 12, 8\}, \{2, 10, 11, 3\}, \{4, 7, 9, 6\}\}$$
$$\mathcal{A}_H = \{\{1, 3, 9\}, \{2, 6, 5\}, \{4, 12, 10\}, \{7, 8, 11\}\}$$

Let

$$\mathbb{F}_{\mathcal{A}_G}[x] = \{f \in \mathbb{F}[x] : f \text{ is constant on } A_i, i = 1, 2, 3;\ \deg f < |\mathbb{F}^*|\}$$
$$\mathbb{F}_{\mathcal{A}_G}[x] = \langle 1, x^4, x^8 \rangle, \quad \mathbb{F}_{\mathcal{A}_H}[x] = \langle 1, x^3, x^6, x^9 \rangle$$

Multiple recovering sets: Example

Take $\mathbb{F} = \mathbb{F}_{13}$; $G, H \le \mathbb{F}^*$; $G = \langle 5 \rangle, H = \langle 3 \rangle$

$$\mathcal{A}_G = \{\{1,5,12,8\},\{2,10,11,3\},\{4,7,9,6\}\}$$
$$\mathcal{A}_H = \{\{1,3,9\},\{2,6,5\},\{4,12,10\},\{7,8,11\}\}$$

Let

$$\mathbb{F}_{\mathcal{A}_G}[x] = \{f \in \mathbb{F}[x] : f \text{ is constant on } A_i, i = 1,2,3;\ \deg f < |\mathbb{F}^*|\}$$
$$\mathbb{F}_{\mathcal{A}_G}[x] = \langle 1, x^4, x^8 \rangle, \quad \mathbb{F}_{\mathcal{A}_H}[x] = \langle 1, x^3, x^6, x^9 \rangle$$

We construct an LRC $(12, 4, \{2,3\})$, distance ≥ 6, code $\mathcal{C} : \mathbb{F}^4 \to \mathbb{F}^{12}$

$$a = (a_0, a_1, a_2, a_3) \mapsto f_a(x) = a_0 + a_1 x + a_2 x^4 + a_3 x^6$$

$$f_a(x) = \sum_{i=0}^{2} f_i(x) x^i, \text{ where } f_0(x) = a_0 + a_2 x^4,\ f_1(x) = a_1,\ f_2(x) = a_3 x^4; f_i \in \mathbb{F}_{\mathcal{A}}[x]$$

$$f_a(x) = \sum_{j=0}^{1} g_j(x) x^j \text{ where } g_0(x) = a_0 + a_3 x^6, g_1(x) = a_1 + a_2 x^3; g_j \in \mathbb{F}_{\mathcal{A}_H}[x]$$

E.g., $f_a(1)$ can be recovered by computing $\delta_1(x), x \in \{5,12,8\}$ OR $\delta_2(x), x \in \{3,9\}$

Multiple recovering sets

General Construction: $A = \{\alpha_1, \dots, \alpha_n\} \subseteq \mathbb{F}, |A| = n$;

$$A = \overbrace{\sqcup_{i\ge 0} R_i^1}^{\mathcal{A}} = \overbrace{\sqcup_{j\ge 0} R_j^2}^{\mathcal{A}'}; \quad |R_i^1| = r+1, |R_j^2| = s+1$$

$$f_a(x) = \sum_{i=0}^{k-1} a_i g_i(x), \quad g_i(x) \in \mathcal{F}_{\mathcal{A}}^r \cap \mathcal{F}_{\mathcal{A}'}^s$$

Evaluation map: $(a_1, \dots, a_k) \overset{\mathcal{C}}{\mapsto} (f_a(\alpha_1), \dots, f_a(\alpha_n))$

Theorem: Assume that the partitions $\mathcal{A}, \mathcal{A}'$ are *orthogonal*. Then

$$Eval(f : f \in \mathcal{F}_{\mathcal{A}}^r \cap \mathcal{F}_{\mathcal{A}'}^s), x \in A$$

gives an $(n, k, \{r, s\})$ LRC code with distance $\ge n - m + 1$, where m is the largest degree in $\mathcal{F}_{\mathcal{A}}^r \cap \mathcal{F}_{\mathcal{A}'}^s$.

Constructing orthogonal partitions

Orthogonal partitions can be obtained from the structure of additive or multiplicative subgroups of $\mathbb{F}_q$

Constructing orthogonal partitions

Orthogonal partitions can be obtained from the structure of additive or multiplicative subgroups of $\mathbb{F}_q$

1. $\mathbb{F}_{13}$; $G, H \leq \mathbb{F}_{13}^*$; $G = \langle 5 \rangle, H = \langle 3 \rangle$; $G \cap H = \text{id}$

$$\mathcal{A}_G = \{\{1, 5, 12, 8\}, \{2, 10, 11, 3\}, \{4, 7, 9, 6\}\}$$
$$\mathcal{A}_H = \{\{1, 3, 9\}, \{2, 6, 5\}, \{4, 12, 10\}, \{7, 8, 11\}\}$$

Constructing orthogonal partitions

Orthogonal partitions can be obtained from the structure of additive or multiplicative subgroups of $\mathbb{F}_q$

1. $\mathbb{F}_{13}$; $G, H \le \mathbb{F}_{13}^*$; $G = \langle 5 \rangle, H = \langle 3 \rangle$; $G \cap H = \mathrm{id}$

$$\mathcal{A}_G = \{\{1, 5, 12, 8\}, \{2, 10, 11, 3\}, \{4, 7, 9, 6\}\}$$
$$\mathcal{A}_H = \{\{1, 3, 9\}, \{2, 6, 5\}, \{4, 12, 10\}, \{7, 8, 11\}\}$$

2. Take $G, H \le \mathbb{F}_q^+$, e.g., $G \cong H \cong (\mathbb{Z}_2)^2$; $\mathbb{F}_{16}^+ = (\mathbb{Z}_2)^2 \times (\mathbb{Z}_2)^2$

Constructing orthogonal partitions

Orthogonal partitions can be obtained from the structure of additive or multiplicative subgroups of $\mathbb{F}_q$

1. $\mathbb{F}_{13}$; $G, H \le \mathbb{F}_{13}^*$; $G = \langle 5 \rangle, H = \langle 3 \rangle$; $G \cap H = \text{id}$

$$\mathcal{A}_G = \{\{1,5,12,8\},\{2,10,11,3\},\{4,7,9,6\}\}$$
$$\mathcal{A}_H = \{\{1,3,9\},\{2,6,5\},\{4,12,10\},\{7,8,11\}\}$$

2. Take $G, H \le \mathbb{F}_q^+$, e.g., $G \cong H \cong (\mathbb{Z}_2)^2$; $\mathbb{F}_{16}^+ = (\mathbb{Z}_2)^2 \times (\mathbb{Z}_2)^2$

$$G = \{0000, 0001, 0010, 0011\} \text{ and } H = \{0000, 0100, 1000, 1100\}$$
$$\mathcal{A}_G = \{\{0,1,\alpha,\alpha^4\},\{\alpha^5,\alpha^{10},\alpha^2,\alpha^8\},\{\alpha^6,\alpha^{13},\alpha^{11},\alpha^{12}\},\{\alpha^7,\alpha^9,\alpha^{14},\alpha^3\}\}$$
$$\mathcal{A}_H = \{\{0,\alpha^2,\alpha^3,\alpha^6\},\{1,\alpha^8,\alpha^{14},\alpha^{13}\},\{\alpha,\alpha^5,\alpha^9,\alpha^{11}\},\{\alpha^4,\alpha^{10},\alpha^7,\alpha^{12}\}\}$$

Constructing orthogonal partitions

Orthogonal partitions can be obtained from the structure of additive or multiplicative subgroups of $\mathbb{F}_q$

1. $\mathbb{F}_{13}$; $G, H \le \mathbb{F}_{13}^*$; $G = \langle 5 \rangle, H = \langle 3 \rangle$; $G \cap H = \mathrm{id}$

$$\mathcal{A}_G = \{\{1, 5, 12, 8\}, \{2, 10, 11, 3\}, \{4, 7, 9, 6\}\}$$
$$\mathcal{A}_H = \{\{1, 3, 9\}, \{2, 6, 5\}, \{4, 12, 10\}, \{7, 8, 11\}\}$$

2. Take $G, H \le \mathbb{F}_q^+$, e.g., $G \cong H \cong (\mathbb{Z}_2)^2$; $\mathbb{F}_{16}^+ = (\mathbb{Z}_2)^2 \times (\mathbb{Z}_2)^2$

$$G = \{0000, 0001, 0010, 0011\} \text{ and } H = \{0000, 0100, 1000, 1100\}$$
$$\mathcal{A}_G = \{\{0, 1, \alpha, \alpha^4\}, \{\alpha^5, \alpha^{10}, \alpha^2, \alpha^8\}, \{\alpha^6, \alpha^{13}, \alpha^{11}, \alpha^{12}\}, \{\alpha^7, \alpha^9, \alpha^{14}, \alpha^3\}\}$$
$$\mathcal{A}_H = \{\{0, \alpha^2, \alpha^3, \alpha^6\}, \{1, \alpha^8, \alpha^{14}, \alpha^{13}\}, \{\alpha, \alpha^5, \alpha^9, \alpha^{11}\}, \{\alpha^4, \alpha^{10}, \alpha^7, \alpha^{12}\}\}$$

Proposition: Two subgroups G, H define orthogonal coset partitions if they intersect trivially: $G \cap H = \mathrm{id}$

Remarks

There are other ways of constructing codes with multiple (e.g., two) recovering sets:

Product codes, Bipartite-graph codes

Remarks

There are other ways of constructing codes with multiple (e.g., two) recovering sets:

Product codes, Bipartite-graph codes

A family of optimal locally recoverable codes, with **I. Tamo**, arXiv:1311.3284 (*IT Trans.*, no. 8, 2014)

Bounds on the parameters

Theorem

Let $\mathcal{C}$ be an (n,k,r,t) LRC code with t disjoint recovering sets of size r. Then the rate of $\mathcal{C}$ satisfies

$$\frac{k}{n} \le \frac{1}{\prod_{j=1}^{t}(1+\frac{1}{jr})} \approx t^{-\frac{1}{r}}$$

The minimum distance of $\mathcal{C}$ is bounded above as follows:

$$d \le n - \sum_{i=0}^{t} \left\lfloor \frac{k-1}{r^i} \right\rfloor . \qquad (Tamo - B, 2014)$$

$$d \le n - k - \left\lceil \frac{t(k-1)+1}{t(r-1)+1} \right\rceil + 2 \qquad (Rawat\ e.a., 2014)$$

Bounds on the parameters

Theorem

Let $\mathcal{C}$ be an (n, k, r, t) LRC code with t disjoint recovering sets of size r. Then the rate of $\mathcal{C}$ satisfies

$$\frac{k}{n} \leq \frac{1}{\prod_{j=1}^{t}(1 + \frac{1}{jr})} \approx t^{-\frac{1}{r}}$$

The minimum distance of $\mathcal{C}$ is bounded above as follows:

$$d \leq n - \sum_{i=0}^{t} \left\lfloor \frac{k-1}{r^i} \right\rfloor. \qquad (Tamo - B, 2014)$$

$$d \leq n - k - \left\lceil \frac{t(k-1)+1}{t(r-1)+1} \right\rceil + 2 \qquad (Rawat\ e.a., 2014)$$

It is likely that these bounds are not final

LRC codes

A block code of length n over $\mathbb{F}_q$ is called LRC with locality r if every symbol of the codeword can be found by accessing some r symbols of the codeword.

LRC codes

A block code of length n over $\mathbb{F}_q$ is called LRC with locality r if every symbol of the codeword can be found by accessing some r symbols of the codeword.

Regenerating codes: Data can be read off from any location

LRC codes

A block code of length n over $\mathbb{F}_q$ is called LRC with locality r if every symbol of the codeword can be found by accessing some r symbols of the codeword.

Regenerating codes: Data can be read off from any location

Last time we constructed RS-type LRC codes with $q \approx n$, distance meeting the Gopalan et al. Singleton bound:

$$d \leq n - k - \left\lceil \frac{k}{r} \right\rceil + 2$$

LRC codes

A block code of length n over $\mathbb{F}_q$ is called LRC with locality r if every symbol of the codeword can be found by accessing some r symbols of the codeword.

Regenerating codes: Data can be read off from any location

Last time we constructed RS-type LRC codes with $q \approx n$, distance meeting the Gopalan et al. Singleton bound:

$$d \leq n - k - \left\lceil \frac{k}{r} \right\rceil + 2$$

Asymptotic GV bound with locality:

$$R \geq \frac{r}{r+1} - h_q(\delta)$$

Extensions

Reed-Solomon codes can be extended in two ways:

- Codes on algebraic curves
- Cyclic codes and subfield subcodes

Algebraic codes

RS codes: $n \le q-1$, $1 \le k \le n$, $d = n-k+1$

Algebraic codes

RS codes: $n \le q-1$, $1 \le k \le n$, $d = n-k+1$

To construct the code, take $A := (P_1, \dots, P_n) \subset \mathbb{F}_q$

Algebraic codes

RS codes: $n \leq q-1$, $1 \leq k \leq n$, $d = n-k+1$

To construct the code, take $A := (P_1, \ldots, P_n) \subset \mathbb{F}_q$

Message $a = (a_1, \ldots, a_k) \in \mathbb{F}_q^k$

Algebraic codes

RS codes: $n \le q-1$, $1 \le k \le n$, $d = n-k+1$

To construct the code, take $A := (P_1, \dots, P_n) \subset \mathbb{F}_q$

Message $a = (a_1, \dots, a_k) \in \mathbb{F}_q^k \quad \to \quad f(x) = \sum_{i=1}^k a_i x^{i-1}$

Algebraic codes

RS codes: $n \le q-1$, $1 \le k \le n$, $d = n-k+1$

To construct the code, take $A := (P_1, \dots, P_n) \subset \mathbb{F}_q$

Message $a = (a_1, \dots, a_k) \in \mathbb{F}_q^k \quad \to \quad f(x) = \sum_{i=1}^k a_i x^{i-1} \quad \to \quad (f(P_1), \dots, f(P_n))$

Algebraic codes

RS codes: $n \le q-1$, $1 \le k \le n$, $d = n-k+1$

To construct the code, take $A := (P_1, \dots, P_n) \subset \mathbb{F}_q$

$$\boxed{\text{Message } a = (a_1, \dots, a_k) \in \mathbb{F}_q^k \quad \to \quad f(x) = \sum_{i=1}^k a_i x^{i-1} \quad \to \quad (f(P_1), \dots, f(P_n))}$$

Message space:

$$\text{span over } \mathbb{F}_q \text{ of } (x^i, i = 0, \dots, k-1)$$

Algebraic codes

RS codes: $n \le q-1$, $1 \le k \le n$, $d = n-k+1$

To construct the code, take $A := (P_1, \dots, P_n) \subset \mathbb{F}_q$

$$\boxed{\text{Message } a = (a_1, \dots, a_k) \in \mathbb{F}_q^k \quad \to \quad f(x) = \textstyle\sum_{i=1}^k a_i x^{i-1} \quad \to \quad (f(P_1), \dots, f(P_n))}$$

Message space:

$$\text{span over } \mathbb{F}_q \text{ of } (x^i, i = 0, \dots, k-1)$$

By construction, $n \le q$

Algebraic codes

RS codes: $n \le q-1$, $1 \le k \le n$, $d = n-k+1$

To construct the code, take $A := (P_1, \dots, P_n) \subset \mathbb{F}_q$

$$\boxed{\text{Message } a = (a_1, \dots, a_k) \in \mathbb{F}_q^k \quad \to \quad f(x) = \textstyle\sum_{i=1}^k a_i x^{i-1} \quad \to \quad (f(P_1), \dots, f(P_n))}$$

Message space:

$$\text{span over } \mathbb{F}_q \text{ of } (x^i, i = 0, \dots, k-1)$$

By construction, $n \le q$ (recall the *MDS conjecture*)

Algebraic codes

RS codes: $n \le q-1$, $1 \le k \le n$, $d = n-k+1$

To construct the code, take $A := (P_1, \dots, P_n) \subset \mathbb{F}_q$

$$\text{Message } a = (a_1, \dots, a_k) \in \mathbb{F}_q^k \quad \to \quad f(x) = \sum_{i=1}^k a_i x^{i-1} \quad \to \quad (f(P_1), \dots, f(P_n))$$

Message space:

$$\text{span over } \mathbb{F}_q \text{ of } (x^i, i = 0, \dots, k-1)$$

By construction, $n \le q$ (recall the *MDS conjecture*)

Longer codes? We need more points P_i

Algebraic codes

RS codes: $n \le q-1$, $1 \le k \le n$, $d = n-k+1$

To construct the code, take $A := (P_1, \dots, P_n) \subset \mathbb{F}_q$

$$\text{Message } a = (a_1, \dots, a_k) \in \mathbb{F}_q^k \quad \to \quad f(x) = \textstyle\sum_{i=1}^k a_i x^{i-1} \quad \to \quad (f(P_1), \dots, f(P_n))$$

Message space:

$$\text{span over } \mathbb{F}_q \text{ of } (x^i, i = 0, \dots, k-1)$$

By construction, $n \le q$ (recall the *MDS conjecture*)

Longer codes? We need more points P_i

Curves to the rescue!

AG codes in error correction

1. Gilbert-Varshamov bound

An $[n, k, d]$ code exists if

$$\sum_{i=0}^{d-2} \binom{n-1}{i} (q-1)^i < q^{n-k}$$

Let $R = k/n, \delta = d/n$, take logs and divide by n:

$$R \geq 1 - h_q(\delta)$$

AG codes in error correction

1. Gilbert-Varshamov bound

An $[n, k, d]$ code exists if

$$\sum_{i=0}^{d-2} \binom{n-1}{i} (q-1)^i < q^{n-k}$$

Let $R = k/n, \delta = d/n$, take logs and divide by n:

$$R \ge 1 - h_q(\delta)$$

2. Tsfasman-Vlăduţ-Zink bound

There exist explicit sequences of codes on algebraic curves with the parameters

$$R \ge 1 - \delta - \frac{1}{\sqrt{q} - 1}$$

RS type codes

Given $A \subset \mathbb{F}$, partition it into $(r+1)$-subsets.

To encode the message $a \in \mathbb{F}^k$, write $\underline{a} = (a_{ij}, i = 0, \dots, r-1; j = 0, ..., \frac{k}{r} - 1)$

RS type codes

Given $A \subset \mathbb{F}$, partition it into $(r+1)$-subsets.

To encode the message $a \in \mathbb{F}^k$, write $\underline{a} = (a_{ij}, i = 0, \dots, r-1; j = 0, ..., \frac{k}{r} - 1)$

$$\underline{a} \to f_a(x) = \sum_{i=0}^{r-1} f_i(x) x^i, \quad \text{where } f_i(x) = \sum_{j=0}^{\frac{k}{r}-1} a_{ij} g(x)^j, \quad i = 0, ..., r-1$$

RS type codes

Given $A \subset \mathbb{F}$, partition it into $(r+1)$-subsets.

To encode the message $a \in \mathbb{F}^k$, write $\underline{a} = (a_{ij}, i = 0, \dots, r-1; j = 0, ..., \frac{k}{r} - 1)$

$$\underline{a} \to f_a(x) = \sum_{i=0}^{r-1} f_i(x) x^i, \quad \text{where } f_i(x) = \sum_{j=0}^{\frac{k}{r}-1} a_{ij} g(x)^j, \quad i = 0, ..., r-1$$

Message space:

$$\text{span}\,(g(x)^j x^i, i = 0, \dots, r-1; j = 0, \dots, \frac{k}{r} - 1)$$

RS type codes

Given $A \subset \mathbb{F}$, partition it into $(r+1)$-subsets.

To encode the message $a \in \mathbb{F}^k$, write $\underline{a} = (a_{ij}, i = 0, \dots, r-1; j = 0, ..., \frac{k}{r} - 1)$

$$\underline{a} \to f_a(x) = \sum_{i=0}^{r-1} f_i(x) x^i, \quad \text{where } f_i(x) = \sum_{j=0}^{\frac{k}{r}-1} a_{ij} g(x)^j, \quad i = 0, ..., r-1$$

Message space:

$$\text{span}\,(g(x)^j x^i, i = 0, \dots, r-1; j = 0, \dots, \frac{k}{r} - 1)$$

Evaluation code $\mathcal{C}$

$$\begin{aligned} Ev : &\mathbb{F}^k \to \mathbb{F}^n \\ &a \mapsto (f_a(P), P \in A) \end{aligned}$$

RS-like codes

RS-like codes

What is the meaning of $g(x)$?

RS-like codes

What is the meaning of $g(x)$?

It does not make sense that the functions are

$$g(x)^j x^i$$

RS-like codes

What is the meaning of $g(x)$?

It does not make sense that the functions are

$$g(x)^j x^i$$

They should really be

$$g(y)^j x^i$$

Geometric interpretation

$$A := \{1,2,3,4,5,6,9,10,12\} \subset \mathbb{F}_{13}$$

$$g(x) : A \to \mathbb{F}_{13}$$

$$x \mapsto x^3$$

$$A = \{A_1 = \{1,3,9\}, A_2 = \{2,6,5\}, A_3 = \{4,12,10\}\}$$

$$g(A_1) = 1, g(A_2) = 8, g(A_3) = 12$$

Geometric interpretation

$$A := \{1, 2, 3, 4, 5, 6, 9, 10, 12\} \subset \mathbb{F}_{13}$$
$$g(x) : A \to \mathbb{F}_{13}$$
$$x \mapsto x^3$$

$$A = \{A_1 = \{1, 3, 9\}, A_2 = \{2, 6, 5\}, A_3 = \{4, 12, 10\}\}$$
$$g(A_1) = 1, g(A_2) = 8, g(A_3) = 12$$

$$g : \mathbb{F}_q \to \mathbb{F}_q$$
$$|g^{-1}(x)| = r + 1 \text{ for every } x \text{ in the image of } g$$

Geometric interpretation

$$A := \{1,2,3,4,5,6,9,10,12\} \subset \mathbb{F}_{13}$$

$$g(x) : A \to \mathbb{F}_{13}$$
$$x \mapsto x^3$$

$$A = \{A_1 = \{1,3,9\}, A_2 = \{2,6,5\}, A_3 = \{4,12,10\}\}$$
$$g(A_1) = 1, g(A_2) = 8, g(A_3) = 12$$

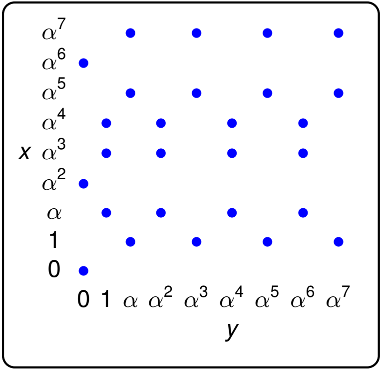
$$g : \mathbb{F}_q \to \mathbb{F}_q$$
$$|g^{-1}(x)| = r + 1 \text{ for every } x \text{ in the image of } g$$

$$\begin{array}{cccc} & 1 & 2 & 4 \\ X & 3 & 6 & 12 \\ & 9 & 5 & 10 \end{array}$$

$$\begin{array}{cccc} Y & 1 & 8 & 12 \end{array}$$

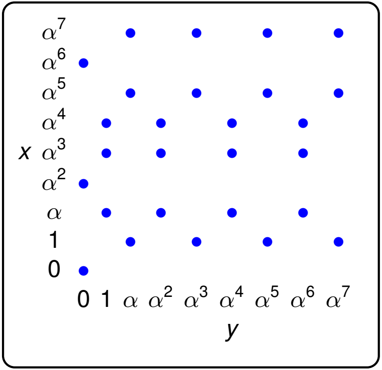
LRC codes on curves

Consider the set of pairs $(x, y) \in \mathbb{F}_9$ that satisfy the equation $x^3 + x = y^4$



LRC codes on curves

Consider the set of pairs $(x, y) \in \mathbb{F}_9$ that satisfy the equation $x^3 + x = y^4$



27 points of the Hermitian curve over $\mathbb{F}_9$; $\alpha^2 = \alpha + 1$

LRC codes on curves

Recall RS codes: $\mathcal{C}$ is a mapping $V_k = \langle 1, x, \ldots, x^{k-1} \rangle \to \mathbb{F}_q^n$

LRC codes on curves

Recall RS codes: $\mathcal{C}$ is a mapping $V_k = \langle 1, x, \ldots, x^{k-1} \rangle \to \mathbb{F}_q^n$

Hermitian codes

Take the space of functions $V := \langle 1, y, y^2, x, xy, xy^2 \rangle$

A={27 points of the Hermitian curve over $\mathbb{F}_9$}; $n = 27, k = 6$

$$\mathcal{C} : V \to \mathbb{F}_9^n$$

LRC codes on curves

Recall RS codes: $\mathcal{C}$ is a mapping $V_k = \langle 1, x, \dots, x^{k-1} \rangle \to \mathbb{F}_q^n$

Hermitian codes

Take the space of functions $V := \langle 1, y, y^2, x, xy, xy^2 \rangle$

A=$\{$27 points of the Hermitian curve over $\mathbb{F}_9\}$; $n = 27, k = 6$

$$\mathcal{C} : V \to \mathbb{F}_9^n$$

E.g., message $(1, \alpha, \alpha^2, \alpha^3, \alpha^4, \alpha^5)$

$$F(x, y) = 1 + \alpha y + \alpha^2 y^2 + \alpha^3 x + \alpha^4 xy + \alpha^5 xy^2$$

$$F(0, 0) = 1 \text{ etc.}$$

LRC codes on curves

x \ y	0	1	α	α^2	α^3	α^4	α^5	α^6	α^7
α^7			α		α^7		α^5		0
α^6	α^2								
α^5			α^6		α^4		α^2		0
α^4		α^7		α^3		α^5		α^5	
α^3		α^3		α^7		α		α	
α^2	α^3								
α		0		0		0		0	
1			1		α^6		α^4		0
0	1								

LRC codes on curves

	α^7			α		α^7		α^5		0
	α^6	α^2								
	α^5			α^6		α^4		α^2		0
	α^4		α^7		α^3		α^5		α^5	
x	α^3		α^3		α^7		α		α	
	α^2	α^3								
	α		$\cancel{0}$		0		0		0	
	1			1		α^6		α^4		0
	0	1								
		0	1	α	α^2	α^3	α^4	α^5	α^6	α^7

y

Let $P = (\alpha, 1)$ be the erased location.

LRC codes on curves

x \ y	0	1	α	α^2	α^3	α^4	α^5	α^6	α^7
α^7			α		α^7		α^5		0
α^6	α^2								
α^5			α^6		α^4		α^2		0
α^4		α^7		α^3		α^5		α^5	
α^3		α^3		α^7		α		α	
α^2	α^3								
α		?		0		0		0	
1			1		α^6		α^4		0
0	1								

Let $P = (\alpha, 1)$ be the erased location. Recovering set $I_P = \{(\alpha^4, 1), (\alpha^3, 1)\}$
Find $f(x) : f(\alpha^4) = \alpha^7, f(\alpha^3) = \alpha^3$

$$\Rightarrow \quad f(x) = \alpha x - \alpha^2$$

LRC codes on curves

x \ y	0	1	α	α^2	α^3	α^4	α^5	α^6	α^7
α^7			α		α^7		α^5		0
α^6	α^2								
α^5			α^6		α^4		α^2		0
α^4		α^7		α^3		α^5		α^5	
α^3		α^3		α^7		α		α	
α^2	α^3								
α		?		0		0		0	
1			1		α^6		α^4		0
0	1								

Let $P = (\alpha, 1)$ be the erased location. Recovering set $I_P = \{(\alpha^4, 1), (\alpha^3, 1)\}$
Find $f(x) : f(\alpha^4) = \alpha^7, f(\alpha^3) = \alpha^3$

$$\Rightarrow \quad f(x) = \alpha x - \alpha^2$$

$$f(\alpha) = 0 = F(P)$$

Hermitian codes

$q = q_0^2$, q_0 prime power

Hermitian codes

$q = q_0^2$, q_0 prime power

$$X : x^{q_0} + x = y^{q_0+1}$$

Hermitian codes

$q = q_0^2$, q_0 prime power

$$X : x^{q_0} + x = y^{q_0+1}$$

X has $q_0^3 = q^{3/2}$ points in $\mathbb{F}_q$

Hermitian codes

$q = q_0^2$, q_0 prime power

$$X : x^{q_0} + x = y^{q_0+1}$$

X has $q_0^3 = q^{3/2}$ points in $\mathbb{F}_q$

Let $g : X \to Y = \mathbb{P}^1$, $g(P) = g(x, y) := y$

Hermitian codes

$q = q_0^2$, q_0 prime power

$$X : x^{q_0} + x = y^{q_0+1}$$

X has $q_0^3 = q^{3/2}$ points in $\mathbb{F}_q$

Let $g : X \to Y = \mathbb{P}^1$, $g(P) = g(x, y) := y$

We obtain a family of q-ary codes of length $n = q_0^3$,

$$k = (t+1)(q_0 - 1), d \geq n - tq_0 - (q_0 - 2)(q_0 + 1)$$

with locality $r = q_0 - 1$.

Hermitian codes

$q = q_0^2$, q_0 prime power

$$X : x^{q_0} + x = y^{q_0+1}$$

X has $q_0^3 = q^{3/2}$ points in $\mathbb{F}_q$

Let $g : X \to Y = \mathbb{P}^1$, $g(P) = g(x, y) := y$

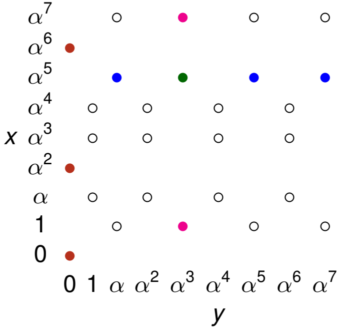
We obtain a family of q-ary codes of length $n = q_0^3$,

$$k = (t+1)(q_0 - 1), d \geq n - tq_0 - (q_0 - 2)(q_0 + 1)$$

with locality $r = q_0 - 1$.

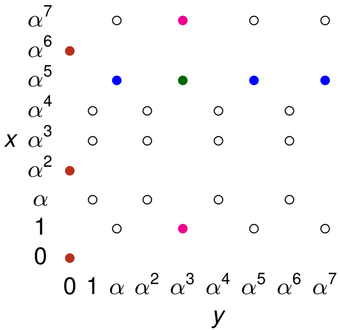
It is also possible to take $g(P) = x$ (projection on x); we obtain LRC codes with locality q_0

Two recovering sets



Polynomial basis $\{x^i y^j, i = 0, 1, \ldots, r_1 - 1, j = 0, 1, \ldots, r_2 - 1\}$

Two recovering sets



Polynomial basis $\{x^i y^j, i = 0, 1, \ldots, r_1 - 1, j = 0, 1, \ldots, r_2 - 1\}$

$(24, 6, \{2, 3\})$ LRC(2) code over $\mathbb{F}_9$

General LRC codes on curves

Map of curves

X, Y smooth projective absolutely irreducible curves over $\Bbbk$

$$g : X \to Y$$

rational separable map of degree $r+1$.

General LRC codes on curves

Map of curves

X, Y smooth projective absolutely irreducible curves over $\mathbb{k}$

$$g : X \to Y$$

rational separable map of degree $r + 1$.

Lift the points of Y

$S = \{P_1, \ldots, P_s\} \subset Y(\mathbb{k})$; $Q_\infty = \pi^{-1}(\infty)$, where $\pi : Y \to \mathbb{P}^1_{\mathbb{k}}$. Assume that there is a partition of points

$$A := g^{-1}(S) = \{P_{ij}, i = 0, \ldots, r, j = 1, \ldots, s\} \subseteq X(\mathbb{k})$$

such that

$$g(P_{ij}) = P_j \text{ for all } i, j.$$

General LRC codes on curves

Map of curves

X, Y smooth projective absolutely irreducible curves over $\mathbb{k}$

$$g : X \to Y$$

rational separable map of degree $r + 1$.

Lift the points of Y

$S = \{P_1, \dots, P_s\} \subset Y(\mathbb{k})$; $Q_\infty = \pi^{-1}(\infty)$, where $\pi : Y \to \mathbb{P}^1_{\mathbb{k}}$. Assume that there is a partition of points

$$A := g^{-1}(S) = \{P_{ij}, i = 0, \dots, r, j = 1, \dots, s\} \subseteq X(\mathbb{k})$$

such that

$$g(P_{ij}) = P_j \text{ for all } i, j.$$

Basis of the function space

$$\{f_j x^i, i = 0, \dots, r-1; j = 1, \dots, m\}$$

General LRC codes on curves

Map of curves

X, Y smooth projective absolutely irreducible curves over $\Bbbk$

$$g : X \to Y$$

rational separable map of degree $r + 1$.

Lift the points of Y

$S = \{P_1, \ldots, P_s\} \subset Y(\Bbbk)$; $Q_\infty = \pi^{-1}(\infty)$, where $\pi : Y \to \mathbb{P}^1_{\Bbbk}$. Assume that there is a partition of points

$$A := g^{-1}(S) = \{P_{ij}, i = 0, \ldots, r, j = 1, \ldots, s\} \subseteq X(\Bbbk)$$

such that

$$g(P_{ij}) = P_j \text{ for all } i, j.$$

Basis of the function space

$$\{f_j x^i, i = 0, \ldots, r - 1; j = 1, \ldots, m\}$$

Construct LRC codes

Evaluation codes constructed on the set A have the locality property with parameter r.

Asymptotically good sequences of codes

Let $q = q_0^2$, where q_0 is a prime power. Take Garcia-Stichtenoth towers of curves:

$$x_0 := 1;\ X_1 := \mathbb{P}^1, \Bbbk(X_1) = \Bbbk(x_1);$$
$$X_l : z_l^{q_0} + z_l = x_{l-1}^{q_0+1}, x_{l-1} := \frac{z_{l-1}}{x_{l-2}} \in \Bbbk(X_{l-1}) \text{ (if } l \geq 3\text{)},$$

Asymptotically good sequences of codes

Let $q = q_0^2$, where q_0 is a prime power. Take Garcia-Stichtenoth towers of curves:

$$x_0 := 1;\ X_1 := \mathbb{P}^1, \Bbbk(X_1) = \Bbbk(x_1);$$
$$X_l : z_l^{q_0} + z_l = x_{l-1}^{q_0+1}, x_{l-1} := \frac{z_{l-1}}{x_{l-2}} \in \Bbbk(X_{l-1}) \text{ (if } l \geq 3),$$

There exist families of q-ary LRC codes with locality r whose *rate and relative distance* satisfy

$$R \geq \frac{r}{r+1}\Big(1 - \delta - \frac{3}{\sqrt{q}+1}\Big), \qquad r = \sqrt{q} - 1$$
$$R \geq \frac{r}{r+1}\Big(1 - \delta - \frac{2\sqrt{q}}{q-1}\Big), \qquad r = \sqrt{q}$$

(better than the GV bound)

Asymptotically good sequences of codes

Let $q = q_0^2$, where q_0 is a prime power. Take Garcia-Stichtenoth towers of curves:

$$x_0 := 1;\ X_1 := \mathbb{P}^1, \Bbbk(X_1) = \Bbbk(x_1);$$
$$X_l : z_l^{q_0} + z_l = x_{l-1}^{q_0+1}, x_{l-1} := \frac{z_{l-1}}{x_{l-2}} \in \Bbbk(X_{l-1}) \text{ (if } l \geq 3),$$

There exist families of q-ary LRC codes with locality r whose *rate and relative distance* satisfy

$$R \geq \frac{r}{r+1}\Big(1 - \delta - \frac{3}{\sqrt{q}+1}\Big), \qquad r = \sqrt{q} - 1$$
$$R \geq \frac{r}{r+1}\Big(1 - \delta - \frac{2\sqrt{q}}{q-1}\Big), \qquad r = \sqrt{q}$$

(better than the GV bound)

$^{*)}$Recall the TVZ bound without locality: $R \geq 1 - \delta - \frac{1}{\sqrt{q}-1}$

Asymptotically good sequences of codes

Let $q=q_0^2$, where q_0 is a prime power. Take Garcia-Stichtenoth towers of curves:

$$x_0 := 1;\ X_1 := \mathbb{P}^1, \mathbb{k}(X_1)=\mathbb{k}(x_1);$$
$$X_l : z_l^{q_0}+z_l = x_{l-1}^{q_0+1}, x_{l-1} := \frac{z_{l-1}}{x_{l-2}} \in \mathbb{k}(X_{l-1}) \text{ (if } l\ge 3),$$

There exist families of q-ary LRC codes with locality r whose *rate and relative distance* satisfy

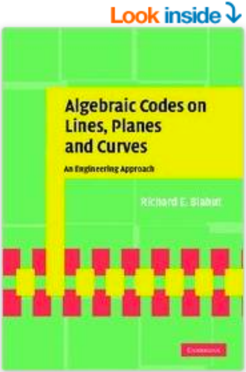
$$R \ge \frac{r}{r+1}\Big(1-\delta-\frac{3}{\sqrt{q}+1}\Big), \qquad r=\sqrt{q}-1$$
$$R \ge \frac{r}{r+1}\Big(1-\delta-\frac{2\sqrt{q}}{q-1}\Big), \qquad r=\sqrt{q}$$

(better than the GV bound)

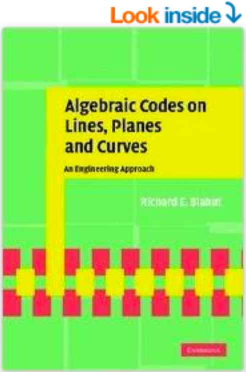
Locally recoverable codes on algebraic curves, with **I. Tamo** and **S. Vlăduţ**, arXiv:1501.04904

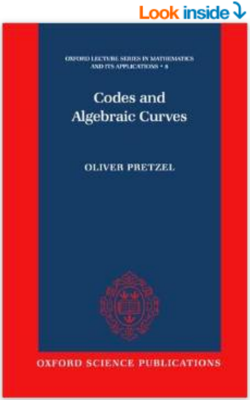
What next?

What next?

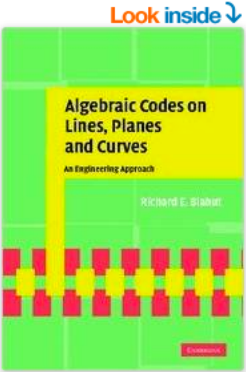

Look inside
Algebraic Codes on
Lines, Planes
and Curves
An Engineering Approach
Richard E. Blahut

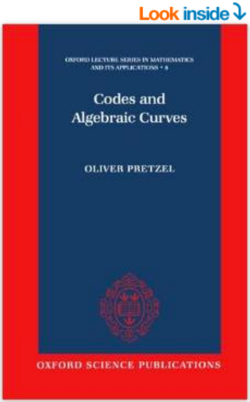
What next?

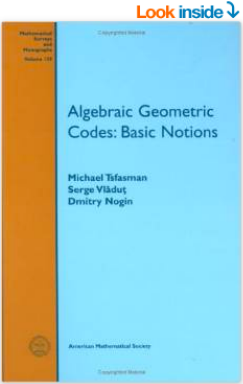

Look inside
Algebraic Codes on
Lines, Planes
and Curves
An Engineering Approach


Look inside
Codes and
Algebraic Curves
OLIVER PRETZEL
OXFORD SCIENCE PUBLICATIONS

What next?


Look inside
Algebraic Codes on Lines, Planes and Curves
An Engineering Approach
Richard E. Blahut


Look inside
Codes and Algebraic Curves
OLIVER PRETZEL
OXFORD SCIENCE PUBLICATIONS


Look inside
Algebraic Geometric Codes: Basic Notions
Michael Tsfasman
Serge Vlăduţ
Dmitry Nogin

Another connection: Cyclic codes and Binary cyclic codes

Another connection: Cyclic codes and Binary cyclic codes

Consider an $[n = 15, k = 4]$ RS code over $\mathbb{F}_{16}$; $A = \{1, \alpha, \alpha^2, \dots, \alpha^{14}\}$

Another connection: Cyclic codes and Binary cyclic codes

Consider an $[n = 15, k = 4]$ RS code over $\mathbb{F}_{16}$; $A = \{1, \alpha, \alpha^2, \ldots, \alpha^{14}\}$

message (a_1, a_2, a_3, a_4); $f(x) = a_1 + a_2x + a_3x^2 + a_4x^3$

$$f(1) = \langle (a_1, a_2, a_3, a_4), (1, 1, 1, 1) \rangle$$

Another connection: Cyclic codes and Binary cyclic codes

Consider an $[n = 15, k = 4]$ RS code over $\mathbb{F}_{16}$; $A = \{1, \alpha, \alpha^2, \dots, \alpha^{14}\}$

message (a_1, a_2, a_3, a_4); $f(x) = a_1 + a_2 x + a_3 x^2 + a_4 x^3$

$$f(1) = \langle (a_1, a_2, a_3, a_4), (1, 1, 1, 1) \rangle$$

$$f(\alpha) = \langle (a_1, a_2, a_3, a_4), (1, \alpha, \alpha^2, \alpha^3) \rangle$$

Another connection: Cyclic codes and Binary cyclic codes

Consider an $[n = 15, k = 4]$ RS code over $\mathbb{F}_{16}$; $A = \{1, \alpha, \alpha^2, \dots, \alpha^{14}\}$

message (a_1, a_2, a_3, a_4); $f(x) = a_1 + a_2 x + a_3 x^2 + a_4 x^3$

$$f(1) = \langle (a_1, a_2, a_3, a_4), (1, 1, 1, 1) \rangle$$

$$f(\alpha) = \langle (a_1, a_2, a_3, a_4), (1, \alpha, \alpha^2, \alpha^3) \rangle$$

$$f(\alpha^2) = \langle (a_1, a_2, a_3, a_4), (1, \alpha^2, \alpha^4, \alpha^6) \rangle$$

Another connection: Cyclic codes and Binary cyclic codes

Consider an $[n = 15, k = 4]$ RS code over $\mathbb{F}_{16}$; $A = \{1, \alpha, \alpha^2, \dots, \alpha^{14}\}$

message (a_1, a_2, a_3, a_4); $f(x) = a_1 + a_2 x + a_3 x^2 + a_4 x^3$

$$f(1) = \langle (a_1, a_2, a_3, a_4), (1, 1, 1, 1) \rangle$$

$$f(\alpha) = \langle (a_1, a_2, a_3, a_4), (1, \alpha, \alpha^2, \alpha^3) \rangle$$

$$f(\alpha^2) = \langle (a_1, a_2, a_3, a_4), (1, \alpha^2, \alpha^4, \alpha^6) \rangle$$

Generator matrix

$$G = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \alpha & \alpha^2 & \dots & \alpha^{14} \\ 1 & \alpha^2 & \alpha^4 & \dots & \alpha^{2\cdot 14} \\ 1 & \alpha^3 & \alpha^6 & \dots & \alpha^{3\cdot 14} \end{pmatrix}$$

Parity-check matrix

Another connection: Cyclic codes and Binary cyclic codes

Consider an $[n = 15, k = 4]$ RS code over $\mathbb{F}_{16}$; $A = \{1, \alpha, \alpha^2, \dots, \alpha^{14}\}$

message (a_1, a_2, a_3, a_4); $f(x) = a_1 + a_2 x + a_3 x^2 + a_4 x^3$

$$f(1) = \langle (a_1, a_2, a_3, a_4), (1, 1, 1, 1) \rangle$$

$$f(\alpha) = \langle (a_1, a_2, a_3, a_4), (1, \alpha, \alpha^2, \alpha^3) \rangle$$

$$f(\alpha^2) = \langle (a_1, a_2, a_3, a_4), (1, \alpha^2, \alpha^4, \alpha^6) \rangle$$

Generator matrix

$$G = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \alpha & \alpha^2 & \dots & \alpha^{14} \\ 1 & \alpha^2 & \alpha^4 & \dots & \alpha^{2\cdot 14} \\ 1 & \alpha^3 & \alpha^6 & \dots & \alpha^{3\cdot 14} \end{pmatrix}$$

Parity-check matrix

$$H = \begin{pmatrix} 1 & \alpha & \alpha^2 & \dots & \alpha^{14} \\ 1 & \alpha^2 & \alpha^{2\cdot 2} & \dots & \alpha^{14\cdot 2} \\ \vdots & & \vdots & & \vdots \\ 1 & \alpha^{11} & \alpha^{2\cdot 11} & \dots & \alpha^{14\cdot 11} \end{pmatrix}$$

Another connection: Cyclic codes and Binary cyclic codes

Consider an $[n = 15, k = 4]$ RS code over $\mathbb{F}_{16}$; $A = \{1, \alpha, \alpha^2, \dots, \alpha^{14}\}$

message (a_1, a_2, a_3, a_4); $f(x) = a_1 + a_2 x + a_3 x^2 + a_4 x^3$

$$f(1) = \langle (a_1, a_2, a_3, a_4), (1, 1, 1, 1) \rangle$$

$$f(\alpha) = \langle (a_1, a_2, a_3, a_4), (1, \alpha, \alpha^2, \alpha^3) \rangle$$

$$f(\alpha^2) = \langle (a_1, a_2, a_3, a_4), (1, \alpha^2, \alpha^4, \alpha^6) \rangle$$

Generator matrix

$$G = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \alpha & \alpha^2 & \dots & \alpha^{14} \\ 1 & \alpha^2 & \alpha^4 & \dots & \alpha^{2\cdot 14} \\ 1 & \alpha^3 & \alpha^6 & \dots & \alpha^{3\cdot 14} \end{pmatrix}$$

Parity-check matrix

$$H = \begin{pmatrix} 1 & \alpha & \alpha^2 & \dots & \alpha^{14} \\ 1 & \alpha^2 & \alpha^{2\cdot 2} & \dots & \alpha^{14\cdot 2} \\ \vdots & & \vdots & & \vdots \\ 1 & \alpha^{11} & \alpha^{2\cdot 11} & \dots & \alpha^{14\cdot 11} \end{pmatrix}$$

$$\underline{c} = (c_1, \dots, c_{15}); \quad c(x) = \sum_{i=1}^{15} c_i x^{i-1} : \quad c(\alpha^i) = 0, i = 1, \dots, 14$$

BCH codes: Subfield subcodes of RS codes

- Consider the subset of vectors of the RS code with coordinates 0 or 1
- $c(x) = \sum_{i=1}^{n} x^i : c(\alpha^j) = 0$
- They form a BCH code, a binary cyclic code of length $2^m - 1$
- This construction is called a Subfield Subcode
  Observation 1: expand parity-check matrix
  Observaion 2: conjugate roots

Cyclic codes

- Consider an $[n|(q-1), k=n-d+1, d]$ RS code $\mathcal{C}$ over $\mathbb{F}_q$

  $A=(1,\alpha,\dots,\alpha^{n-1})$ where $\alpha^n=1$

  Generator matrix

  $$G=\begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \alpha & \alpha^2 & \dots & \alpha^{n-1} \\ \vdots & & \vdots & & \vdots \\ 1 & \alpha^{k-1} & \alpha^{2(k-1)} & \dots & \alpha^{(k-1)(n-1)} \end{pmatrix}$$

  Parity-check matrix

  $$H=\begin{pmatrix} 1 & \alpha & \alpha^2 & \dots & \alpha^{n-1} \\ 1 & \alpha^2 & \alpha^{2\cdot 2} & \dots & \alpha^{2(n-1)} \\ \vdots & & \vdots & & \vdots \\ 1 & \alpha^{n-k} & \alpha^{2(n-k)} & \dots & \alpha^{(n-k)(n-1)} \end{pmatrix}$$

  $\mathcal{C}$ is a cyclic code with zeros $\alpha,\alpha^2,\dots,\alpha^{n-k}$

Cyclic codes

- Consider an $[n|(q-1), k=n-d+1, d]$ RS code $\mathcal{C}$ over $\mathbb{F}_q$

  $A=(1,\alpha,\dots,\alpha^{n-1})$ where $\alpha^n=1$

  Generator matrix

  $$G=\begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \alpha & \alpha^2 & \dots & \alpha^{n-1} \\ \vdots & & \vdots & & \vdots \\ 1 & \alpha^{k-1} & \alpha^{2(k-1)} & \dots & \alpha^{(k-1)(n-1)} \end{pmatrix}$$

  Parity-check matrix

  $$H=\begin{pmatrix} 1 & \alpha & \alpha^2 & \dots & \alpha^{n-1} \\ 1 & \alpha^2 & \alpha^{2\cdot 2} & \dots & \alpha^{2(n-1)} \\ \vdots & & \vdots & & \vdots \\ 1 & \alpha^{n-k} & \alpha^{2(n-k)} & \dots & \alpha^{(n-k)(n-1)} \end{pmatrix}$$

  $\mathcal{C}$ is a cyclic code with zeros $\alpha,\alpha^2,\dots,\alpha^{n-k}$

- Consider a subfield subcode $\mathcal{D}\subset\mathcal{C}$,
  $\mathcal{D}:=\{(c_0,\dots,c_{n-1})\in\mathcal{C} : c_j\in\mathbb{F}_p, 0\le j\le n-1\}$
  Zeros of $\mathcal{D}$: $\{(\alpha,\alpha^p,\dots,\alpha^{p^{m-1}\mathrm{mod} n}),\dots\}$

Cyclic codes: Example

- RS code $\mathcal{C}$ of length $n = 15, k = 8, d = 8, q = 2^4$
  Zeros of $\mathcal{C}$: $\alpha, \alpha^2, \alpha^3, \alpha^4, \alpha^5, \alpha^6, \alpha^7$
  Generator polynomial $g(x) = \prod_{i=1}^{t}(x - \alpha^i)$, $\dim(\mathcal{C}) = n - \deg(g) = 8$

BCH bound: $d(\mathcal{C}) \geq$ number of consecutive 0's $+ 1$

Cyclic codes: Example

- RS code $\mathcal{C}$ of length $n = 15, k = 8, d = 8, q = 2^4$
  Zeros of $\mathcal{C}$: $\alpha, \alpha^2, \alpha^3, \alpha^4, \alpha^5, \alpha^6, \alpha^7$
  Generator polynomial $g(x) = \prod_{i=1}^{t}(x - \alpha^i)$, $\dim(\mathcal{C}) = n - \deg(g) = 8$

BCH bound: $d(\mathcal{C}) \geq$ number of consecutive 0's $+ 1$

- Now suppose that $\mathcal{C}$ has zeros $\{\alpha, \dots, \alpha^7\} \cup \{\alpha, \alpha^4, \alpha^7, \alpha^{10}, \alpha^{13}\}$. The distance is the same, and we get the locality condition $r = 2$

Cyclic codes: Example

- RS code $\mathcal{C}$ of length $n = 15, k = 8, d = 8, q = 2^4$
  Zeros of $\mathcal{C}$: $\alpha, \alpha^2, \alpha^3, \alpha^4, \alpha^5, \alpha^6, \alpha^7$
  Generator polynomial $g(x) = \prod_{i=1}^{t}(x - \alpha^i)$, $\dim(\mathcal{C}) = n - \deg(g) = 8$

BCH bound: $d(\mathcal{C}) \geq$ number of consecutive 0's $+ 1$

- Now suppose that $\mathcal{C}$ has zeros $\{\alpha, \dots, \alpha^7\} \cup \{\alpha, \alpha^4, \alpha^7, \alpha^{10}, \alpha^{13}\}$. The distance is the same, and we get the locality condition $r = 2$
  Indeed,

$$f_a(x) = a_1 + a_2 x + a_3 x^3 + a_4 x^4 + a_5 x^6 + a_6 x^7$$

Cyclic codes: Example

- RS code $\mathcal{C}$ of length $n = 15, k = 8, d = 8, q = 2^4$
  Zeros of $\mathcal{C}$: $\alpha, \alpha^2, \alpha^3, \alpha^4, \alpha^5, \alpha^6, \alpha^7$
  Generator polynomial $g(x) = \prod_{i=1}^{t}(x - \alpha^i)$, $\dim(\mathcal{C}) = n - \deg(g) = 8$

BCH bound: $d(\mathcal{C}) \geq$ number of consecutive 0's $+ 1$

- Now suppose that $\mathcal{C}$ has zeros $\{\alpha, \dots, \alpha^7\} \cup \{\alpha, \alpha^4, \alpha^7, \alpha^{10}, \alpha^{13}\}$. The distance is the same, and we get the locality condition $r = 2$
  Indeed,

$$f_a(x) = a_1 + a_2 x + a_3 x^3 + a_4 x^4 + a_5 x^6 + a_6 x^7$$

So the rows of G are $1, \alpha, \alpha^3, \alpha^4, \alpha^6, \alpha^7$

Cyclic codes: Example

- RS code $\mathcal{C}$ of length $n = 15, k = 8, d = 8, q = 2^4$
  Zeros of $\mathcal{C}$: $\alpha, \alpha^2, \alpha^3, \alpha^4, \alpha^5, \alpha^6, \alpha^7$
  Generator polynomial $g(x) = \prod_{i=1}^{t}(x - \alpha^i)$, $\dim(\mathcal{C}) = n - \deg(g) = 8$

BCH bound: $d(\mathcal{C}) \geq$ number of consecutive 0's $+ 1$

- Now suppose that $\mathcal{C}$ has zeros $\{\alpha, \ldots, \alpha^7\} \cup \{\alpha, \alpha^4, \alpha^7, \alpha^{10}, \alpha^{13}\}$. The distance is the same, and we get the locality condition $r = 2$
  Indeed,

$$f_a(x) = a_1 + a_2 x + a_3 x^3 + a_4 x^4 + a_5 x^6 + a_6 x^7$$

So the rows of G are $1, \alpha, \alpha^3, \alpha^4, \alpha^6, \alpha^7$
The rows of H are $-([n]\backslash\{0, 1, 3, 4, 6, 7\})$

Cyclic codes: Example

- RS code $\mathcal{C}$ of length $n = 15, k = 8, d = 8, q = 2^4$
  Zeros of $\mathcal{C}$: $\alpha, \alpha^2, \alpha^3, \alpha^4, \alpha^5, \alpha^6, \alpha^7$
  Generator polynomial $g(x) = \prod_{i=1}^{t}(x - \alpha^i)$, $\dim(\mathcal{C}) = n - \deg(g) = 8$

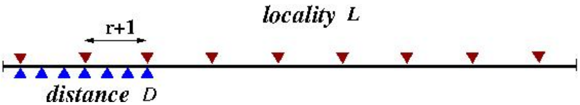
BCH bound: $d(\mathcal{C}) \geq$ number of consecutive 0's $+ 1$

- Now suppose that $\mathcal{C}$ has zeros $\{\alpha, \dots, \alpha^7\} \cup \{\alpha, \alpha^4, \alpha^7, \alpha^{10}, \alpha^{13}\}$. The distance is the same, and we get the locality condition $r = 2$
  Indeed,
  $$f_a(x) = a_1 + a_2 x + a_3 x^3 + a_4 x^4 + a_5 x^6 + a_6 x^7$$
  So the rows of G are $1, \alpha, \alpha^3, \alpha^4, \alpha^6, \alpha^7$
  The rows of H are $-([n] \backslash \{0, 1, 3, 4, 6, 7\})$

$$k = 6; d = 8 = n - k\frac{r+1}{r} + 2$$

Cyclic LRC codes

Main idea: Suppose that the zeros are arranged as follows:



The cyclic code with zeros $\{D \cup L\}$ has distance $\geq |D|$ and locality r.

Cyclic LRC codes: Details

The following result describes the cyclic case of the main construction.

Theorem (RS-type cyclic LRC codes): Let α be a primitive n-th root of unity, where $n|(q-1)$; let $l, 0 \le l \le r$ be an integer. Consider the following sets of elements of $\mathbb{F}_q$:

$$L = \{\alpha^i, i \bmod (r+1) = l\},$$

and

$$D = \Big\{\alpha^{j+s}, s = 0, \dots, n - \frac{k}{r}(r+1)\Big\},$$

where $\alpha^j \in L$. The cyclic code with the defining set of zeros $L \cup D$ is an optimal $^{*)}$ (n, k, r) q-ary cyclic LRC code.

$^{*)}$ Singleton-like optimality; see (1)

Locality and dual distance

Let $\mathcal{C}$ be a cyclic LRC code over $\mathbb{F}_q$.

Locality and dual distance

Let $\mathcal{C}$ be a cyclic LRC code over $\mathbb{F}_q$.

Dual code

$$\mathcal{C}^\perp := \{\underline{x} \in \mathbb{F}_q : \langle \underline{x}, \underline{c} \rangle = 0 \text{ for all } \underline{c} \in \mathcal{C}\}$$

Locality and dual distance

Let $\mathcal{C}$ be a cyclic LRC code over $\mathbb{F}_q$.

Dual code

$$\mathcal{C}^\perp := \{\underline{x} \in \mathbb{F}_q : \langle \underline{x}, \underline{c} \rangle = 0 \text{ for all } \underline{c} \in \mathcal{C}\}$$

Locality of $\mathcal{C}$:

$$r = d(\mathcal{C}^\perp) = d^\perp(\mathcal{C})$$

In the cyclic case Locality=Dual distance

Subfield subcodes

What about binary codes?

Subfield subcodes

What about binary codes?

Example:
Code $\mathcal{C}$ over $\mathbb{F}_{16}$ has zeros $Z = \{\alpha, \alpha^2, \alpha^3, \alpha^4\} \cup \{\alpha, \alpha^4, \alpha^7, \alpha^{10}, \alpha^{13}\}$.

Subfield subcodes

What about binary codes?

Example:
Code $\mathcal{C}$ over $\mathbb{F}_{16}$ has zeros $Z = \{\alpha, \alpha^2, \alpha^3, \alpha^4\} \cup \{\alpha, \alpha^4, \alpha^7, \alpha^{10}, \alpha^{13}\}$.

Binary subcode $\mathcal{D} \subset \mathcal{C}$: zeros Z and all conjugates
The locality of D may decrease; the distance may increase. The dimension becomes smaller.

Subfield subcodes

What about binary codes?

Subfield subcodes

What about binary codes?

Let $\mathcal{C}$ be a cyclic code over $\mathbb{F}_{q^m}$; let $\mathcal{D}$ be the subfield subcode of $\mathcal{C}$

$$\mathcal{D} := \{\underline{c} = (c_1, \dots, c_n) \in \mathcal{C} : c_i \in \mathbb{F}_q, i = 1, \dots, n\}$$

Subfield subcodes

What about binary codes?

Let $\mathcal{C}$ be a cyclic code over $\mathbb{F}_{q^m}$; let $\mathcal{D}$ be the subfield subcode of $\mathcal{C}$

$$\mathcal{D} := \{\underline{c} = (c_1, \dots, c_n) \in \mathcal{C} : c_i \in \mathbb{F}_q, i = 1, \dots, n\}$$

We have:

$$d(\mathcal{D}) \geq d(\mathcal{C})$$

Subfield subcodes

What about binary codes?

Let $\mathcal{C}$ be a cyclic code over $\mathbb{F}_{q^m}$; let $\mathcal{D}$ be the subfield subcode of $\mathcal{C}$

$$\mathcal{D} := \{\underline{c} = (c_1, \dots, c_n) \in \mathcal{C} : c_i \in \mathbb{F}_q, i = 1, \dots, n\}$$

We have:

$$d(\mathcal{D}) \geq d(\mathcal{C})$$

$$d^{\perp}(\mathcal{D}) \leq d^{\perp}(\mathcal{C})$$

Subfield subcodes

What about binary codes?

Let $\mathcal{C}$ be a cyclic code over $\mathbb{F}_{q^m}$; let $\mathcal{D}$ be the subfield subcode of $\mathcal{C}$

$$\mathcal{D} := \{\underline{c} = (c_1, \dots, c_n) \in \mathcal{C} : c_i \in \mathbb{F}_q, i = 1, \dots, n\}$$

We have:

$$d(\mathcal{D}) \geq d(\mathcal{C})$$

$$d^{\perp}(\mathcal{D}) \leq d^{\perp}(\mathcal{C})$$

$$r(\mathcal{D}) \leq r(\mathcal{C})$$

Subfield subcodes

The analysis: Ideas.

- Take a subfield subcode D of the code $\mathcal{C}$ constructed in the RS-like LRC codes Theorem.
- Locality of D = distance of $D^{\perp}$
- Let $q = 2^m$, $T_m(x) = x + x^2 + \cdots + x^{2^{m-1}}, x \in \mathbb{F}_q$

$$T_m(\mathcal{C}) := \{(T_m(c_1), \ldots, T_m(c_n)), \underline{c} \in \mathcal{C}\}$$

Theorem (Delsarte '74, Sidelnikov '71): $D = T_m(\mathcal{C}^{\perp})$

- Analyze the locality of D using $d(D^{\perp})$ (techniques: irreducible cyclic codes)

Some examples

n	k	d	$Z(\mathcal{D})$	r	w	$Z((\mathcal{C}')^{\perp})$	$d^{\perp}$	SH	LP	locator field
35	20	3	$\{1, 15\}$	$r \le 3$	4	$\{0, 1, 7, 15\}$	4	$k \le 25$	$k \le 29$	$\mathbb{F}_{2^{12}}$
45	33	3	$\{1\}$	$r \le 7$	8	$\{0, 1, 3, 5, 9, 15, 21\}$	8	$k \le 37$	$k \le 39$	$\mathbb{F}_{2^{12}}$
27	7	6	$\{1, 9\}$	$r = 1$	2	$\{0, 3\}$	2			$\mathbb{F}_{2^{18}}$
63	36	3	$\{1, 9, 11, 15, 23\}$	$r \le 3$	4	$\{0, 1, 7, 9, 11, 15, 21, 23\}$	4			$\mathbb{F}_{2^{6}}$

$Z(\mathcal{C})$ =defining set of of zeros of $\mathcal{C}$, w is the number of recovering sets A_i

Some examples

n	k	d	$Z(\mathcal{D})$	r	w	$Z((\mathcal{C}')^\perp)$	$d^\perp$	SH	LP	locator field
35	20	3	$\{1, 15\}$	$r \le 3$	4	$\{0, 1, 7, 15\}$	4	$k \le 25$	$k \le 29$	$\mathbb{F}_{2^{12}}$
45	33	3	$\{1\}$	$r \le 7$	8	$\{0, 1, 3, 5, 9, 15, 21\}$	8	$k \le 37$	$k \le 39$	$\mathbb{F}_{2^{12}}$
27	7	6	$\{1, 9\}$	$r = 1$	2	$\{0, 3\}$	2			$\mathbb{F}_{2^{18}}$
63	36	3	$\{1, 9, 11, 15, 23\}$	$r \le 3$	4	$\{0, 1, 7, 9, 11, 15, 21, 23\}$	4			$\mathbb{F}_{2^6}$

$Z(\mathcal{C})$ =defining set of of zeros of $\mathcal{C}$, w is the number of recovering sets A_i

Cyclic LRC codes and their subfield subcodes, with **I. Tamo, S. Goparaju,** and **R. Calderbank**, arXiv:1502.01414.