

# Topics in Algebra: Cryptography

**Univ.-Prof. Dr. Goulmara ARZHANTSEVA**

WS 2018



# Code

Let  $\mathcal{P}$  be a finite set of possible messages.

## Definition: Code

A **code** is a subset  $C \subset \{0, 1\}^n$  with  $|C| = |\mathcal{P}|$ , and an **encoding** is given by a **bijjective** map  $\psi: \mathcal{P} \rightarrow C$ .

# Code

Let  $\mathcal{P}$  be a finite set of possible messages.

## Definition: Code

A **code** is a subset  $C \subset \{0, 1\}^n$  with  $|C| = |\mathcal{P}|$ , and an **encoding** is given by a **bijjective** map  $\psi: \mathcal{P} \rightarrow C$ .

## Linear code

A **linear code**  $C$  is a code with  $\mathcal{P} = \{0, 1\}^k$  for some  $k < n$ , and encoding is done by a linear operator (a **generating matrix**)

$$A_C \in \mathbb{F}_2^{k \times n}:$$

$$\psi(v) = v^T A_C.$$

A linear code  $C$  is a linear subspace of  $\{0, 1\}^n$  of **dimension**  $k$ , whose basis is given by the rows of  $k \times n$  matrix  $A_C$ .

# Distance and rate of a code

For  $x, y \in C$ , the **Hamming distance**  $d_H(x, y)$  = number of distinct bits

Definition: Distance and rate

$$d(C) = \min_{\substack{x, y \in C \\ x \neq y}} d_H(x, y), \quad r(C) = \frac{\log_2 |C|}{n}$$

# Distance and rate of a code

For  $x, y \in C$ , the **Hamming distance**  $d_H(x, y)$  = number of distinct bits

Definition: Distance and rate

$$d(C) = \min_{\substack{x, y \in C \\ x \neq y}} d_H(x, y), \quad r(C) = \frac{\log_2 |C|}{n}$$

The distance measures the ability to resolve corrupted bits.

The **distance should be large**: two codewords should be sufficiently dissimilar so that corruption of a single bit (or of a small number of bits) does not turn one codeword into another.

## What code will correct $t$ -bit errors?

If 2 bits are bad in a codeword  $a$ , the resulting (erroneous) codeword  $a'$  is at distance  $2 = d_H(a, a')$ .

Such errors can be corrected if  $d(a', c) \geq 3$ : the correct codeword  $a$  is the closest to  $a'$ . Thus,  $d(C) \geq 5$  is required:

$$5 \leq d(C) \leq d_H(a, c) \leq d_H(a, a') + d_H(a', c).$$

Similarly, we obtain the following result.

### Observation

A code with  $d(C) \geq t + (t + 1) = 2t + 1$  can correct  $t$ -bit errors.

## Distance and rate of a code

For  $x, y \in C$ , the **Hamming distance**  $d_H(x, y)$  = number of distinct bits

Definition: Distance and rate

$$d(C) = \min_{\substack{x, y \in C \\ x \neq y}} d_H(x, y), \quad r(C) = \frac{\log_2 |C|}{n}$$

The rate measures the number of information-bits.

For a linear code of dimension  $k$ , the rate is  $\frac{\log_2 2^k}{n} = \frac{k}{n}$ , the amount of non-redundant information per bit.

The **rate should be large**.

# Distance versus rate

A sparser code has larger distance (i.e. more errors can be corrected) but smaller rate (i.e. smaller information-density).

Theorem: Quantifying the distance-rate tradeoff

Hamming'1950

Let  $C \subset \{0, 1\}^n$  be a code and  $t = \lfloor \frac{d(C)-1}{2} \rfloor$ . Then

$$\frac{|C|}{2^n} \leq \frac{1}{\sum_{i=0}^t \binom{n}{i}}$$

A code is **perfect** if it achieves the Hamming bound.



## Distance versus rate

**Proof:** For  $x \in \{0, 1\}^n$ ,  $B(x, t) = \{y \in \{0, 1\}^n \mid d_H(x, y) \leq t\}$  is the ball of radius  $t$  centered at  $x$ , with respect to the Hamming distance.

For all  $x, y \in C$ ,  $x \neq y$ , the sets  $B(x, t)$  and  $B(y, t)$  are disjoint.

Otherwise,  $d_H(x, y) \leq 2t < d(C)$ , contradicting the definition of  $d(C)$ .

Each  $B(x, t)$  has size  $\sum_{i=0}^t \binom{n}{i}$ . Their union is contained in  $\{0, 1\}^n$ , so

$$|C| \sum_{i=0}^t \binom{n}{i} \leq 2^n$$



# Distance versus rate

## Example: the Hamming bound

A linear code of length  $n$ , dimension  $k$  and distance 3 satisfies

$$k \leq n - \log_2(n + 1)$$

## Example: a Hamming code of length 7, dimension 4 and distance 3

For  $(x_1, x_2, x_3, x_4) \in \{0, 1\}^4$ , we define

$$C_{\text{Ham}}(x_1, x_2, x_3, x_4) = (x_1, x_2, x_3, x_4, x_2 \oplus x_3 \oplus x_4, x_1 \oplus x_3 \oplus x_4, x_1 \oplus x_2 \oplus x_4)$$

$4 = 7 - \log_2(7 + 1)$ , hence,  $C_{\text{Ham}}$  has the largest possible dimension for any binary code of length 7 and distance 3.

## Distance of a linear code

For  $c \in \mathbb{F}_2^n$ , the support  $\text{supp } c$  = the set of positions with nonzero bits.

$|\text{supp } c|$  = the number of nonzero bits, the **weight** of  $c$ .

### Lemma: Distance of a linear code

For a linear code  $C$ , we have  $d(C) = \min_{c \in C, c \neq 0} |\text{supp } c|$ .

**Proof:** If  $c, c' \in C$ , then  $c \oplus c' \in C$ , since  $C$  is linear. Then,

$$d_H(c, c') = d_H(0, c \oplus c') = |\text{supp } (c \oplus c')|.$$



# Asymptotically good codes

## Definition: Asymptotically good code

A family  $\mathcal{C}$  of codes  $C_n \subset \{0, 1\}^n$  as  $n \rightarrow \infty$ , is **asymptotically good** if there exist constants  $\alpha, \lambda > 0$  such that for all  $C_n \in \mathcal{C}$ ,

$$\frac{d(C_n)}{n} > \alpha \text{ and } r(C_n) > \lambda.$$

We want both a constant-fraction number of errors and a constant rate.

We also want that encoding and decoding is **in P**, ideally in **linear time**.

# Bipartite graphs

## Definition: A bipartite graph

A graph is **bipartite** if there is a partition of its set of vertices into two (disjoint) subsets  $S$  and  $T$  such that every edge has one endpoint vertex in  $S$  and another one in  $T$ .

## Definition: An $(l, r)$ -regular graph

A bipartite graph is  $(l, r)$ -regular if all vertices in  $S$  have degree  $l$ , and all vertices in  $T$  have degree  $r$ .

The complete graph  $K_{3,3}$  is a bipartite graph. It is  $(3, 3)$ -regular.

# Bipartite expander graphs

## Definition: Bipartite expander

A bipartite graph  $X$  is an  $(l, r, \alpha, \delta)$ -**expander** if it is  $(l, r)$ -regular, and for all sets  $U \subset S$  with  $|U| \leq \alpha|S|$ , we have  $|\partial U| > \delta|U|$ .

$\partial U$  denotes the **external boundary** of the set  $U$  = the set of vertices at distance 1 from  $U$ , in the edge-length distance on  $X$ .

Here,  $\alpha, \delta > 0$  are real constants and  $l, r$  are positive integers.

Small subsets of  $S$  have big enough boundary: they are ‘**expanding**’.

# Parity check matrix

Let  $S \subset \mathbb{F}_2^r$  be an  $r$ -bit linear code of dim.  $k$  with **parity check matrix**  $P_S$ :

$$c \in S \iff P_S c = 0.$$

The  $(r - k) \times r$  matrix  $P_S$  describes linear relations that hold  $\forall c \in S$ .

Rows of  $A_S$  span  $S$  and rows of  $P_S$  span  $S^\perp$ .

That is,  $P_S A_S^T = (0)$ , the zero matrix of size  $(r - k) \times k$ .

# Parity check matrix

Let  $S \subset \mathbb{F}_2^r$  be an  $r$ -bit linear code of dim.  $k$  with **parity check matrix**  $P_S$ :

$$c \in S \iff P_S c = 0.$$

The  $(r - k) \times r$  matrix  $P_S$  describes linear relations that hold  $\forall c \in S$ .

Rows of  $A_S$  span  $S$  and rows of  $P_S$  span  $S^\perp$ .

That is,  $P_S A_S^T = (0)$ , the zero matrix of size  $(r - k) \times k$ .

By a changing of basis of  $\mathbb{F}_2^r$ , we write  $A_S = (I_k M)$ , where  $I_k$  the  $k \times k$  identity matrix.

Then,  $P_S = (M^T I_{r-k})$ .



## Towards expander codes

Let  $X$  be an  $(l, r)$ -regular expander whose  $l$ -degree side has  $n$  vertices and  $l < r$ .

We will extend an  $r$ -bit linear code  $S$  to an  $n$ -bit linear code  $C(X, S)$ .

This will allow to produce an asymptotically good family of codes.

# Expander codes

Let  $\{u_1, \dots, u_n\}$  be  $n$  vertices on the  $l$ -degree side of  $X$ .

Then the  $r$ -degree side has  $(l \cdot n)/r$  vertices, say  $\{v_1, \dots, v_{ln/r}\}$ .

Let  $\sigma$  be a function such that for  $i = 1, \dots, ln/r$ , the neighbours of  $v_i$  are  $u_{\sigma(i,1)}, \dots, u_{\sigma(i,r)}$

**Definition:**  $C(X, S)$

$$C(X, S) = \{(x_1, \dots, x_n) \in \mathbb{F}_2^n \mid \forall i \text{ we have } (x_{\sigma(i,1)}, \dots, x_{\sigma(i,r)}) \in S\}$$

# Expander codes

Lemma: Expander code is linear

$C(X, S)$  is a linear code.

**Proof:** If  $B_{X,i}$  is the 0 – 1 matrix that maps  $(x_1, \dots, x_n)$  to  $(x_{\sigma(i,1)}, \dots, x_{\sigma(i,r)})$ , of size  $r \times n$ , then the parity check matrix  $P_{C(X,S)}$  is the matrix whose rows are the union of the rows of the matrices  $P_S B_{X,i}$ , each of size  $(r - k) \times n$ , for  $i = 1, \dots, \lfloor n/r \rfloor$ . ■

# Expander code

Theorem: Expander code

Sipser-Spielman'1994

Suppose that  $X$  is an  $(l, r, \alpha, l/r\epsilon)$ -expander, and  $S$  has rate  $R > 1 - 1/l$  and normalised distance  $d(S)/r = \epsilon$ . Then  $C(X, S)$  has rate at least  $1 - l(1 - R)$  and normalised distance at least  $\alpha$ .

# Expander code

## Theorem: Expander code

Sipser-Spielman'1994

Suppose that  $X$  is an  $(l, r, \alpha, l/r\epsilon)$ -expander, and  $S$  has rate  $R > 1 - 1/l$  and normalised distance  $d(S)/r = \epsilon$ . Then  $C(X, S)$  has rate at least  $1 - l(1 - R)$  and normalised distance at least  $\alpha$ .

**Proof:** Each matrix  $P_S B_{X,i}$  has  $r - k = (1 - R) \cdot r$  rows. So, the parity check matrix of  $C(X, S)$  has

$$\frac{l \cdot n}{r}(1 - R)r = ln(1 - R) \text{ rows.}$$

This spans  $C(X, S)^\perp$ . Hence, the dimension of  $C(X, S)$  is at least  $n - ln(1 - R)$ , and rate of at least  $\frac{n - ln(1 - R)}{n} = 1 - l(1 - R)$ .

Next we bound the normalised distance  $d(C(X, S))/n$ .

## Expander code: Theorem (suite)

Suppose by contradiction that there is  $c \in C(X, S)$  with  $|\text{supp } c| \leq \alpha n$ .

Let  $U$  be the vertices in  $X$  corresponding to the coordinates of  $\text{supp } c$ .

By the expansion of the graph,  $|\partial U| > \frac{1}{r\epsilon} |U|$ . There are  $|U|$  edges from  $U$  to  $X \setminus U$ , so some  $v_i \in \partial U$  has  $< r\epsilon$  neighbours in  $U$ .

Then,  $(x_{\sigma(i,1)}, \dots, x_{\sigma(i,r)}) \in S$  has  $< r\epsilon$  1-bits, contradicting the hypothesis that the normalised distance of  $S$  is  $\epsilon$ .



# Asymptotically good error-correcting codes

## Corollary

If  $(X_i)_{i \geq 1}$  is a family of  $(l, r, \alpha, l/r\epsilon)$ -expanders with  $n$  vertices of its  $l$ -degree side, as  $n \rightarrow \infty$ , then  $C(X_i, S)_{i \geq 1}$  are asymptotically good error-correcting codes.

# Expander code: Example

## Expander code from the even-weight code

Let  $S_{\text{even}} \subset \mathbb{F}_2^r$  be the code consisting of all even-weight codewords. Then  $P_{S_{\text{even}}} = (1 \ 1 \ \dots \ 1)$ , the normalised distance of  $S_{\text{even}}$  is  $2/r$  and the rate  $R = 1 - 1/r$ .

If  $X$  is an  $(l, r, \alpha, l/2)$ -expander, then, by the Theorem,  $C(X, S_{\text{even}})$  has the normalised distance at least  $\alpha$  and the rate at least  $1 - l/r$ .



# Linear error-correcting (without proof)

Theorem: Linear decoding

Sipser-Spielman'1994

If  $X$  is an  $(l, r, \alpha, \frac{3}{4}l)$ -expander, then the code  $C(X, S_{\text{even}})$  permits an  $\alpha/2$  fraction of errors to be corrected in linear time.

$C(X, S_{\text{even}})$  has normalised distance at least  $\alpha$  and rate at least  $1 - l/r$ .

# Linear error-correcting

There is a linear-time algorithm that will map to a codeword any word of relative distance at most  $\alpha$  from that codeword, for some positive constant  $\alpha$ .

## Algorithm

While not all constraints are satisfied, find a variable  $x_i$  in more unsatisfied than satisfied constraints, and switch  $x_i$ .

$C(X, \mathcal{S}_{\text{even}})$  has  $n$  variables and  $(l \cdot n)/r$  constraints.

Given  $(x_1, \dots, x_n) \in \{0, 1\}^n$ , a constraint  $v_i$  is satisfied if  $(x_{\sigma(i,1)}, \dots, x_{\sigma(i,r)}) \in \mathcal{S}_{\text{even}}$ , i.e. the mod2 sum of the coordinates is zero. Otherwise, it is unsatisfied.

# Linear error-correcting

There is a linear-time algorithm that will map to a codeword any word of relative distance at most  $\alpha$  from that codeword, for some positive constant  $\alpha$ .

## Algorithm

While not all constraints are satisfied, find a variable  $x_i$  in more unsatisfied than satisfied constraints, and switch  $x_i$ .

$C(X, \mathcal{S}_{\text{even}})$  has  $n$  variables and  $(l \cdot n)/r$  constraints.

Given  $(x_1, \dots, x_n) \in \{0, 1\}^n$ , a constraint  $v_i$  is satisfied if  $(x_{\sigma(i,1)}, \dots, x_{\sigma(i,r)}) \in \mathcal{S}_{\text{even}}$ , i.e. the mod2 sum of the coordinates is zero. Otherwise, it is unsatisfied.

One shows that the algorithm terminates after linear number of switches and can be implemented in linear time.

# Asymptotically good linear time error-correcting codes

## Corollary

If  $(X_i)_{i \geq 1}$  is a family of  $(l, r, \alpha, \frac{3}{4}l)$ -expanders with  $n$  vertices of its  $l$ -degree side, as  $n \rightarrow \infty$ , then  $C(X_i, S_{\text{even}})_{i \geq 1}$  are asymptotically good linear time error-correcting codes.

# Existence and constructions of expanders: Remarks

Theorem: Existence of expanders

Kolmogorov-Barzdin'1968

A random (bipartite) graph is an expander.

The above definition of expander can be adapted to usual (not necessarily) bipartite graphs.

Examples of explicit (non bipartite) expanders can be produced by taking **box spaces of finitely generated residually finite groups with Kazhdan's property (T)**.

$SL_3(\mathbb{Z})$  is such a group and  $SL_3(\mathbb{Z}/p\mathbb{Z})$  as prime  $p \rightarrow \infty$  is such an (explicit) expander.

# Existence and constructions of expanders: Remarks

A usual expander gives a bipartite expander: take two copies of the vertex set for each finite graph and have an edge between vertices in different copies if and only if there is an edge between these vertices in the original graph.

Expander graphs are ubiquitous in mathematics and computer science!

# Test questions

## Question 27

Is the Hamming distance indeed a distance?

## Question 28

Given a linear code  $C$ , is its generating matrix uniquely defined?

## Question 29

Is the complete graph  $K_{3,3}$  a bipartite expander?

## Question 30

Let  $Y$  is a non bipartite expander with the expansion parameter  $\lambda$ . What is the expansion parameter of the bipartite expander  $X$  constructed from  $Y$  as in the previous slide. What about the diameter and the girth of  $X$  (given the diameter and the girth of  $Y$ )?