Lecture 2: Linear codes. Hamming codes.

Invited lecturer: Grigory Kabatiansky

g.kabatyansky@skoltech.ru

Teaching Assistant: Stanislav Kruglik

stanislav.kruglik@skolkovotech.ru

February 1, 2017

Outline

Linear code

Definition

A subgroup of Abelian group $\mathbb{F}_q^n$ is called a group code.

Definition

A subspace $\mathcal{C}$ of a vector space $\mathbb{F}_q^n$ is called a linear (n, k) code, where $k = \dim \mathcal{C}$.

Definition

G is a generator matrix for the code $\mathcal{C}$ if the rows of G form a basis in $\mathcal{C}$.

Minimum distance

Lemma

Let $\mathcal{C}$ be a linear code, then

$$d(\mathcal{C}) = \min_{a \in \mathcal{C}, a \neq 0} ||a||.$$

Dual code and parity check matrix

Definition (Dual code)

$$C^{\perp} = \{v \in \mathbb{F}_q^n : v \perp \mathcal{C}\}.$$

Definition (Parity check matrix)

H is a basis of $C^{\perp}$.

Varshamov–Gilbert bound

Theorem

If

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

then the exists a linear $(n, k, \geq d)_q$ *code.*

Proof

Proof.

Construct a parity check matrix column by column. Assume we have already added $\hat{n}$ columns and want to add one more column. As any $d-1$ columns are linerly independent (because of the distance) we can not add this number of columns (linear combinations of $d-2$ columns)

$$\sum_{j=0}^{d-2} \binom{\hat{n}}{j} (q-1)^j,$$

If this number is less, than the total number of columns (q^{n-k}) we can continue the procedure.

□

Practically all linear codes are good!

We randomly generate each bit of the parity-check matrix of size $(n-k) \times n$ according to a Bernoulli distribution Bern$(1/2)$. Let us consider a fixed word $\mathbf{x}$ of length n and weight $W > 0$. The probability of this word to be a codeword (or a probability of the syndrome to be equal to zero) can be calculated as follows

$$\Pr(\mathbf{x} = \mathbf{0}) = 2^{-(n-k)}.$$

Indeed, let us find any non-zero bit (say, bit i) in $\mathbf{x}$. Choose all the elements (except the column i) in arbitrarily. The probability to choose the i-th column such, that the syndrome is equal to zero is $2^{-(n-k)}$.

Practically all linear codes are good!

Now consider the following event E: the code includes at least one codeword with weight $W \leq \delta n$. We have

$$\Pr(E) \leq \sum_{i=1}^{\delta n} \binom{n}{i} 2^{-(n-k)} \quad \text{(union bound)}.$$

Finally,

$$\sum_{i=1}^{\delta n} \binom{n}{i} 2^{-(n-k)} \leq 2^{-n(1-R-h(\delta))}$$

and we see, that the probability (or fraction) of bad codes decrease exponentially with n for any $R < 1 - h(\delta)$.

Exercise, 2 points

Prove

Lemma (Bassalygo's lemma)

Let $L \subset \mathbb{F}_q^n$, $A_q^{(L)}(n,d)$ is the maximal number of words with distance d in L, then

$$\frac{A_q(n,d)}{q^n} \leq \frac{A_q^{(L)}(n,d)}{|L|}$$

Outline

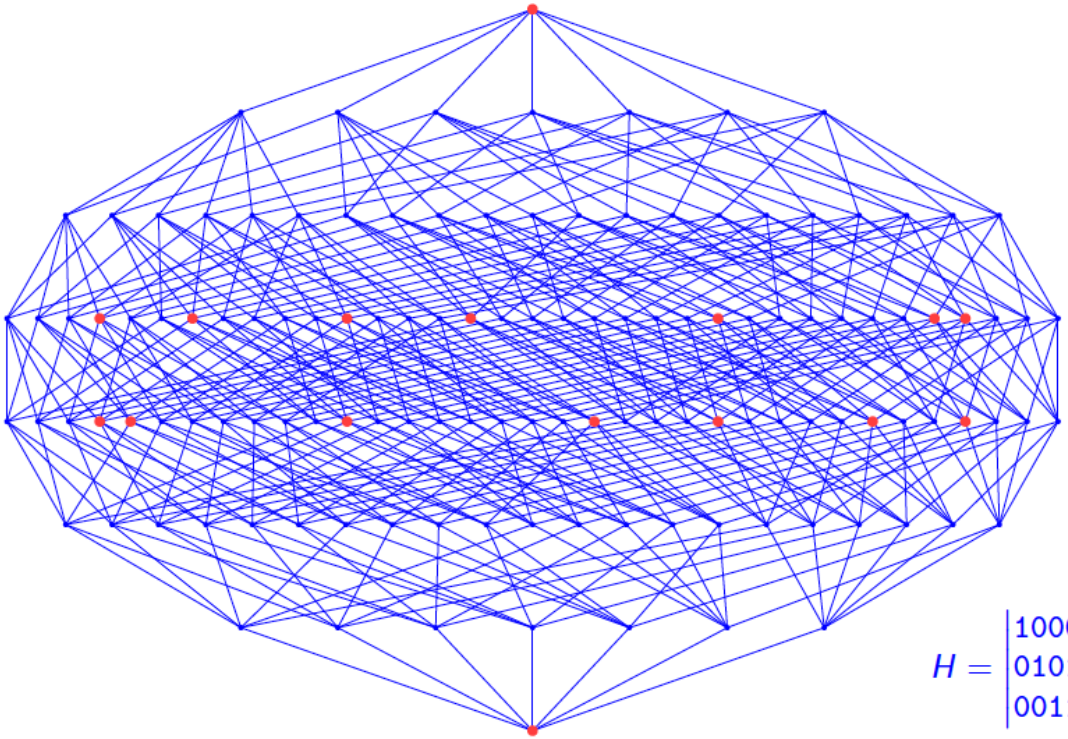
Hamming code

Hamming code is determined by its parity check matrix (PCM) which consists of all non-zero column vectors. Any two columns of such code are linearly independent and there exist 3 linearly dependent columns

Parameters:

$$n = 2^m - 1, k = 2^m - m - 1, d = 3.$$

Hamming code



$$H = \begin{vmatrix} 1000111 \\ 0101011 \\ 0011101 \end{vmatrix}$$

Hamming code

$$H_2 = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}, \quad H_3 = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix},$$

$$H_4 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

Ulam's problem

Choose arbitrary number from 0 to $2^{20} - 1$ and try to determine it with minimal number of "yes-no" questions.

What is the minimal number of questions? The answer is 20 questions. Let us ask a questions of form "does the number belong to a subset". This number of questions is obtained if we take rows of Hamming code PCM as characteristic vectors of such subsets.

And what if the number of questions if some of the answers may be wrong? – 25 – 26 questions

Syndrome calculation

$$\begin{pmatrix} S_1 \\ \vdots \\ S_m \end{pmatrix} = H_m \begin{pmatrix} y_1 \\ \vdots \\ y_{2^m-1} \end{pmatrix} = \left[\begin{array}{c|c|c} 0\,0\,0\,0\,\ldots\,0 & 1 & 1\,1\,1\,1\,\ldots\,1 \\ \hline H_{m-1} & \begin{matrix} 0 \\ 0 \\ \vdots \\ 0 \end{matrix} & H_{m-1} \end{array}\right] \begin{pmatrix} y_1 \\ \vdots \\ y_{2^m-1} \end{pmatrix} =$$

$$= \begin{pmatrix} y_{2^{m-1}} \oplus y_{2^{m-1}+1} \oplus \ldots \oplus y_{2^m-1} \\ H_{m-1} \begin{pmatrix} y_1 \\ \vdots \\ y_{2^{m-1}-1} \end{pmatrix} \oplus H_{m-1} \begin{pmatrix} y_{2^{m-1}+1} \\ \vdots \\ y_{2^m-1} \end{pmatrix} \end{pmatrix} =$$

$$= \begin{pmatrix} y_{2^{m-1}} \oplus y_{2^{m-1}+1} \oplus \ldots \oplus y_{2^m-1} \\ H_{m-1} \begin{pmatrix} y_1 \oplus y_{2^{m-1}+1} \\ \ldots \\ y_{2^{m-1}-1} \oplus y_{2^m-1} \end{pmatrix} \end{pmatrix}$$

Hamming code can correct 1 error. The syndrome is a column of the PCM, which corresponds to error.

Hamming bound and Hamming code

$$n = 2^m - 1$$

$$d = 3$$

$$|\mathcal{C}| = 2^{n-m} = \frac{2^n}{2^m} = \frac{2^n}{n+1}$$

Hamming code lies on the Hamming bound. Such codes are called perfect codes. Another example of perfect code is Golay codes with parameters $n = 23$, $k = 12$, $d = 7$

Outline

Problem 1

Find the number of binary linear $[n, k]$ codes.

Problem 2

Let us have generator matrix of a binary code in form

$$\mathbf{G} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}, \tag{1}$$

- Find its' systematic form
- Write a set of all codewords of this code
- Find it's PCM in systematic form
- Find all parity-check equations
- Find the code distance

Problem 3

Find the number of words of weight 3 and 4 in the binary Hamming code of lenght $n = 2^m - 1$.

Problem 4

Let G_1 and G_2 be the generator matrices of linear $[n_1, k, d_1]$ and $[n_2, k, d_2]$ codes. Find the parameters of code with generator matrix

$$\mathrm{G}= \begin{bmatrix} \mathrm{G}_1 & 0 \\ 0 & \mathrm{G}_2 \end{bmatrix} (2)$$

Problem 5

Using the binary linear code with PCM

$$\mathbf{H} = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}, \tag{3}$$

decode the received vector $v = (111101)$.

Problem 6

Prove that for any integer $r > 1$ a $[\frac{q^r-1}{q-1}, \frac{q^r-1}{q-1} - r, 3]_q$ is Hamming code

Thank you for your attention!