MTAT.05.082: Introduction to Coding Theory **University of Tartu**

Final exam

Instructor: Dr. Vitaly Skachek June 10th, 2014

Student name: ______________________________

Student ID: ______________________________

1. This exam contains 10 pages. Check that no pages are missing.
2. It is possible to collect up to 110 points. Try to collect as many points as possible.
3. Justify and prove all your answers.
4. All facts and results that were proved or stated in the class can be used in your solution without a proof. Such results need to be rigorously formulated.
5. Any printed and written material is allowed in the class. No electronic devices are allowed.
6. Exam duration is 3 hours.
7. Good luck!

Question 1	
Question 2	
Question 3	
Question 4	
Total	

Question 1 (15 points). Let $\mathcal{C}$ be an $[n, k, d]$ code over $\mathbb{F}_2$, defined by the following parity-check matrix:

$$H = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 \end{pmatrix} .$$

What are the values of n, k and d? Justify your answer.

Solution

- It is straightforward to see that the length of the code is $n = 6$.
- The first three rows of H are linearly independent (for example, the 3×3 submatrix formed by the first three rows and the first three columns of H is identity). The fourth row of H is obtained as a sum of the first and the third row. Therefore, the rank of H is 3. We have

$$n - k = \text{rank}(H) = 3 \qquad \Longrightarrow \qquad k = 3 .$$

- It was shown in the class that the minimum distance of the code is the largest number d such that any $d-1$ columns in H are linearly independent. For the given H, any two columns are linearly independent (since they are all different). However, there are three columns which are linearly dependent (take, for example, columns 1, 2 and 4). Therefore, the minimum distance is $d = 3$.

Question 2 (25 points).

Definition: let w and n be integers, $0 \le w \le n$. A code $\mathcal{C}$ is called a *code of constant weight w and of length n over* $\mathbb{F}_2$ if all vectors in $\mathcal{C}$ are binary vectors of length n having Hamming weight w.

Example: for $w = 3$ and $n = 4$, a code that is formed by *any nonempty subset* of the following set of vectors is a constant weight code:

$$\{(0111), (1011), (1101), (1110)\} \ .$$

Let $\mathcal{C}$ be a code of constant weight w and of length n over $\mathbb{F}_2$.

(a) What is the maximal possible size of $\mathcal{C}$?

(b) Show that for any two codewords $\bar{\boldsymbol{x}}$ and $\bar{\boldsymbol{y}}$ in $\mathcal{C}$, the Hamming distance $\mathsf{d}_H(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}})$ is even.

(c) Define a sphere of radius $r > 0$ around the codeword $\bar{\boldsymbol{x}} \in \mathcal{C}$ as

$$\mathbb{S}_r(\bar{\boldsymbol{x}}) = \{\bar{\boldsymbol{y}} \ : \ \mathsf{w}_H(\bar{\boldsymbol{y}}) = w \text{ and } \mathsf{d}_H(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}) \le r\} \ ,$$

where $\mathsf{w}_H(\bar{\boldsymbol{y}})$ denotes the Hamming weight of $\bar{\boldsymbol{y}}$. What is the size of $\mathbb{S}_r(\bar{\boldsymbol{x}})$?

(d) Let d be an integer, $1 \le d \le n$. By using the results in (a) and (c), formulate and prove an upper bound on the size of a code $\mathcal{C}$ of constant weight w and of length n over $\mathbb{F}_2$ with an additional property that for any two codewords $\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}} \in \mathcal{C}$, $\mathsf{d}_H(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}) \ge d$.

Solution

(a) The total number of binary words of weight w and of length n is given by the binomial coefficient $\binom{n}{w}$.

(b) Let $\bar{\boldsymbol{x}} = (x_1, x_2, \cdots, x_n) \in \mathcal{C}$ and $\bar{\boldsymbol{y}} = (y_1, y_2, \cdots, y_n) \in \mathcal{C}$. Define the word $\bar{\boldsymbol{z}} = \bar{\boldsymbol{x}} + \bar{\boldsymbol{y}}$, where $\bar{\boldsymbol{z}} = (z_1, z_2, \cdots, z_n) \in (\mathbb{F}_2)^n$, and the sum is taken over $\mathbb{F}_2$. Observe that $z_i = 0$ if and only if $x_i = y_i$. Therefore, $\mathsf{d}_H(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}})$ is equal to the Hamming weight of $\bar{\boldsymbol{z}}$.

Let ℓ be a number of coordinates i such that $x_i = y_i = 1$, let m be the number of coordinates such that $x_i = 0$ and $y_i = 1$, and let s be the number of coordinates such that $x_i = 1$ and $y_i = 0$. The number of ones in $\bar{\boldsymbol{z}}$ is $m + s$. On the other hand, the number of ones in $\bar{\boldsymbol{x}}$ and $\bar{\boldsymbol{y}}$ together is $2\ell + m + s$ and it is even. Therefore, $m + s$ is even.

(c) Let $\bar{\boldsymbol{x}} \in \mathcal{C}$ be a center of $\mathbb{S}_r(\bar{\boldsymbol{x}})$ for some integer $r > 0$. The number of words of weight w at distance $2j$ from $\bar{\boldsymbol{x}}$, for $j = 0, 1, 2, \cdots$, is given by

$$\binom{w}{j} \cdot \binom{n-w}{j} ,$$

where we denote that $\binom{k}{j} = 0$ if $k < j$. In this product, the first binomial coefficient gives a number of ways to convert j ones in $\bar{\boldsymbol{x}}$ into zeros, and the second binomial coefficient gives a number of ways to convert j zeros in $\bar{\boldsymbol{x}}$ into ones.

Due to (b), the answer is

$$\sum_{j=0}^{\lfloor r/2 \rfloor} \binom{w}{j} \cdot \binom{n-w}{j} .$$

(d) The same idea as in the sphere-packing bound. Consider the spheres of radius $\lfloor (d-1)/2 \rfloor$ around the codewords. Due to triangle inequality of the Hamming distance, these spheres are disjoint. The total number of words of weight w is computed in (a). The volume of each sphere of radius r is computed in (c). We obtain that

$$|\mathcal{C}| \leq \frac{\binom{n}{w}}{\sum_{j=0}^{\left\lfloor \frac{\lfloor (d-1)/2 \rfloor}{2} \right\rfloor} \binom{w}{j} \cdot \binom{n-w}{j}} .$$

Question 3 (35 points). Let $\mathbb{F}$ be a finite field with q elements, and let $\mathcal{C}$ be an $[n,k,d]$ Reed-Solomon code over $\mathbb{F}$, $q > n$.

(a) Let $\alpha_1, \alpha_2, \cdots, \alpha_k$ be elements in $\mathbb{F}$, and let $i_1, i_2, \cdots, i_k$ be a subset of k different indices from $\{1, 2, \cdots, n\}$. Show that there exists a *unique* codeword $\bar{\boldsymbol{c}} = (c_1, c_2, \cdots, c_n)$ in $\mathcal{C}$ such that $c_{i_1} = \alpha_1$, $c_{i_2} = \alpha_2$, $\ldots$, $c_{i_k} = \alpha_k$.

(b) Let t be an integer, $1 \le t \le k$. Let $\alpha_1, \alpha_2, \cdots, \alpha_t$ be elements in $\mathbb{F}$, and let $i_1, i_2, \cdots, i_t$ be a subset of t different indices from $\{1, 2, \cdots, n\}$. How many codewords $\bar{\boldsymbol{c}} = (c_1, c_2, \cdots, c_n)$ in $\mathcal{C}$ satisfy that $c_{i_1} = \alpha_1$, $c_{i_2} = \alpha_2$, $\ldots$, $c_{i_t} = \alpha_t$? Justify your answer.

(c) How many codewords in $\mathcal{C}$ have zeros in the first $k-1$ coordinates?

(d) How many codewords of Hamming weight d does the code $\mathcal{C}$ have?

Solution

(a) Let H be an $(n-k) \times n$ parity-check matrix of $\mathcal{C}$, and $\bar{\boldsymbol{c}}$ be a codeword in $\mathcal{C}$. Denote $\bar{\boldsymbol{c}} = (c_1, c_2, \cdots, c_n)$. Set $c_{i_1} = \alpha_1$, $c_{i_2} = \alpha_2$, $\ldots$, $c_{i_k} = \alpha_k$. Denote by S the set of coordinates

$$S = \{1, 2, \cdots, n\} \backslash \{i_1, i_2, \cdots, i_k\} .$$

In other words, S is the set of all the coordinates in $\bar{\boldsymbol{c}}$, which are not fixed yet. Obviously, $|S| = n - k$.

It is known that $\bar{\boldsymbol{c}} \in \mathcal{C}$ if and only if $H \cdot \bar{\boldsymbol{c}}^T = \mathbf{0}^T$. Since k coordinates of $\bar{\boldsymbol{c}}$ are fixed, this is equivalent to

$$H' \cdot \bar{\boldsymbol{b}}^T = \bar{\boldsymbol{\gamma}}^T ,$$

where H' is an $(n-k) \times (n-k)$ submatrix of H defined by the columns in the set S, $\bar{\boldsymbol{b}}$ is a subvector of $\bar{\boldsymbol{c}}$ restricted to the coordinates in S, and $\bar{\boldsymbol{\gamma}} \in \mathbb{F}^{n-k}$ is the vector of length $n-k$. The matrix H' is a Vandermonde matrix, – it was shown in the class that it is invertible. We multiply by $(H')^{-1}$ from the left, and obtain that

$$\bar{\boldsymbol{b}}^T = (H')^{-1} \cdot H' \cdot \bar{\boldsymbol{b}}^T = (H')^{-1} \cdot \bar{\boldsymbol{\gamma}}^T$$

is a unique solution for $\bar{\boldsymbol{b}}$ as required. This $\bar{\boldsymbol{b}}$ corresponds to a unique solution for $\bar{\boldsymbol{c}}$.

(b) Denote by P the set of coordinates

$$P = \{1, 2, \cdots, n\} \backslash \{i_1, i_2, \cdots, i_t\} .$$

In other words, P is the set of all the coordinates in $\bar{\boldsymbol{c}}$, which are not fixed yet. Obviously, $|P| = n - t$.

Take a set R of $k-t$ arbitrary coordinates in P. Denote $S = P \backslash R$. The set S is of cardinality $(n-t) - (k-t) = n - k$. For any entry c_i of $\bar{\boldsymbol{c}}$, for $i \in R$, there are q ways to establish its value. Then, from (a) with respect to this set S, for any selection of values for coordinates in R, there exists a unique codeword.

There are q^{k-t} ways to select values for entries indexed by coordinates in R. Each selection gives one codeword in $\mathcal{C}$. Therefore, there are q^{k-t} codewords in $\mathcal{C}$ as required.

(c) This is a special case of (b), when $t = k - 1$, $\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$, and $i_1 = 1, i_2 = 2, \cdots, i_{k-1} = k - 1$. From (b), there are $q^{k-(k-1)} = q$ codewords.

(d) Select arbitrary $k - 1$ coordinates $i_1, i_2, \cdots, i_{k-1}$ to be zeros, and select a coordinate i_k to satisfy $c_{i_k} = \alpha \in \mathbb{F}$. For this selection, from (a), there exist a unique codeword.

If $\alpha = 0$, then this codeword has $k = n - d + 1$ zeros, and therefore it has strictly less than d nonzeros. It must be all-zero codeword. If $\alpha \neq 0$, since the codeword has Hamming weight at least d, the remaining $n - k = d - 1$ coordinates all must be nonzeros, and then the total Hamming weight of the word is d. We conclude that for this selection of $i_1, i_2, \cdots, i_{k-1}$, there are $q - 1$ codewords of weight d.

We obtain that the total number of codewords of Hamming weight d is

$$\binom{n}{k-1} \cdot (q - 1) \, ,$$

where the binomial coefficient gives the locations of zero-coordinates and $q - 1$ gives the nonzero value of the k-th selected coordinate.

Question 4 (35 points).

Let $\mathbb{F} = \mathbb{F}_8$ be an extension field of $\mathbb{F}_2$ constructed using irreducible polynomial $x^3 + x + 1$, and let β be a primitive element in $\mathbb{F}$. Suppose that $\mathcal{C}$ is a $[5, 3, 3]$ Reed-Solomon code over $\mathbb{F}$, with a parity-check matrix given by

$$H = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & \beta & \beta^2 & \beta^3 & \beta^4 \end{pmatrix} .$$

Assume that $\bar{\boldsymbol{c}} \in \mathbb{F}^5$ is transmitted, and $\bar{\boldsymbol{y}} = (\beta, \beta^3, \beta^2, \beta^5, 1) \in \mathbb{F}^5$ is received.

1. Find the syndrome polynomial $s(x)$.
2. Find the error-locator and the error-evaluator polynomials. Show all intermediate steps in your algorithm.
3. What are the locations and the values of the errors?
4. What is $\bar{\boldsymbol{c}}$ if there was at most one error?

Solution

1. We have
$$H \cdot \bar{\boldsymbol{y}}^T = \begin{pmatrix} \beta^3 \\ \beta^4 \end{pmatrix} .$$
Therefore, $s(x) = \beta^4 x + \beta^3$.

2. Apply the extended Euclid's algorithm with $r_{-1}(x) = x^2$, $r_0(x) = s(x)$, $t_{-1}(x) = 0$ and $t_0(x) = 1$.
From
$$r_{-1}(x) = q_1(x) r_0(x) + r_1(x) \ ,$$
we obtain that $q_1(x) = \beta^3 x + \beta^2$ and $r_1(x) = \beta^5$. In particular, the stopping condition holds: $\deg(r_1(x)) = 0 < (d-1)/2$.
From
$$t_{-1}(x) = q_1(x) t_0(x) + t_1(x) \ ,$$
we obtain that $t_1(x) = \beta^3 x + \beta^2$.
   - Error-locator polynomial is $\Lambda(x) = t_1(x) = \beta^3 x + \beta^2$.
   - Error-evaluator polynomials is $\Gamma(x) = r_1(x) = \beta^5$.

3. By using Chien's search, the root of $\Lambda(x)$ is β^6. Therefore, the corresponding error locator is $(\beta^6)^{-1} = \beta$, and so the error happened in the second coordinate.
By using Forney's algorithm, the corresponding error value is
$$e_2 = -\frac{\beta}{1} \cdot \frac{\Gamma(\beta^6)}{\Lambda'(\beta^6)} = -\frac{\beta \cdot \beta^5}{\beta^6} = \beta^3 \ .$$
The value of the error in coordinate 2 is β^3.

4. We have:
   - $\bar{\boldsymbol{y}} = (\beta, \beta^3, \beta^2, \beta^5, 1)$ is received.
   - $\bar{\boldsymbol{e}} = (0, \beta^3, 0, 0, 0)$ is the error vector.
   - $\bar{\boldsymbol{c}} = (\beta, 0, \beta^2, \beta^5, 1)$ is the transmitted codeword.