Algorithms of Information Security: Error-correcting codes. Tutorial.

Olha Jurečková, Martin Jureček
{jurecolh,jurecmar}@fit.cvut.cz

Faculty of Information Technology
Czech Technical University in Prague

October 13, 2022

Basic definition

Definition

The Hamming distance $d(x, y)$ of two vectors x and y is equal to the number of coordinates in which they differ.

Example. $d(1000111, 1010110) = 2$.

Definition

The generator matrix of a linear $[n, k]$ code C in F^n is a $k \times n$ matrix G, with elements in F, such that its rows form the bases of C.

The matrix G is in the standard form if $G = (I_k \mid A)$, where I_k is the identity $k \times k$ matrix and A is any $k \times (n-k)$ matrix.
The generator matrix has dimension $k \times n$ and must satisfy 3 basic rules:

1. each row of the matrix is a codeword
2. the rows of the matrix are linearly independent, so the rank of the matrix G is equal to k
3. each codeword is a linear combination of matrix rows.

If code C has a generator matrix $G = (I_k \mid A)$, then its control matrix corresponds to $H = (-A^T \mid I_{n-k})$, where I_{n-k} is identity matrix $(n-k) \times (n-k)$.

Linear codes

Example 1. Consider the field F_3 and let the generator matrix of [5,3]-code be as follows:

$$G = \begin{pmatrix} 1 & 0 & 0 & 1 & 2 \\ 0 & 2 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

Convert the matrix G to the standard form and find the parity check matrix H of the code.

Solution: We have the matrix

$$G = \begin{pmatrix} 1 & 0 & 0 & 1 & 2 \\ 0 & 2 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

and we multiply the second row of the matrix by 2 and get the following matrix:

$$G' = \begin{pmatrix} 1 & 0 & 0 & 1 & 2 \\ 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

Linear codes

Then $G' = (I_3 \mid A)$, so the matrix in standard form is

$$A = \begin{pmatrix} 1 & 2 \\ 0 & 2 \\ 1 & 0 \end{pmatrix}$$

Note. The parity check matrix of the linear code C is the generator matrix of its dual code.
The parity check matrix is

$$H = (-A^T \mid I_2) = \begin{pmatrix} 2 & 0 & 2 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \end{pmatrix}.$$

Cyclic codes

Definition

A cyclic code is a linear code whose generator matrix is made up of codewords (vectors).These code words will be generated by a cyclic shift. The linear code C of length n over the field F_q is therefore invariant with respect to the cyclic shift of its coordinates.

For each word $a = (a_0, \dots, a_{n-1}) \in F_q^n$ holds: $(a_0, \dots, a_{n-1}) \in C \Rightarrow (a_1, \dots, a_{n-1}, a_0) \in C$. Each word (vector) a can be identified with a polynomial over the field F_q, i.e., $a = (a_0, \dots, a_{n-1})$ is represented by $a(x) = a_{n-1}x^{n-1} + \dots + a_1 x + a_0$ or

$$a(x) = \sum_{i=0}^{n-1} a_i x^i \in F_q^n[x].$$

Cyclic codes

The polynomials of the polynomial code are then multiples of the generator polynomial since the cyclic shift corresponds to multiplication by the polynomial x. Generator matrix of the cyclic code with the polynomial $a(x) = a_{n-1}x^{n-1} + \ldots + a_1 x + a_0$ is:

$$G = \begin{pmatrix} a_0 & a_1 & \ldots & a_{n-1} & 0 & \ldots & 0 \\ 0 & a_0 & a_1 & \ldots & a_{n-1} & \ldots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & & \\ 0 & 0 & \ldots & a_0 & a_1 & \ldots & a_{n-1} \end{pmatrix}$$

Cyclic codes

Example 2.
Find the generator matrix for the cyclic code (6,3) whose generator polynomial is as follows: $1 \cdot x^3 + 0 \cdot x^2 + 1 \cdot x + 1$.

Solution. Immediately, from the knowledge of the coefficients of the polynomial $x^3 + x + 1$, we get the generator matrix by shifting as follows:

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

Cyclic codes

Example 3. Find the generator and parity check matrix over the field F_2 for the binary cyclic code of length 6 with the generator polynomial: $g(x) = x^3 + 1$.

Solution. We have $n = 6$. Note that we have defined k such that $deg(g(x)) = n - k$, then $n - k = 3$ and hence $k = 3$. The generator matrix is obtained immediately from the knowledge of the coefficients of the polynomial $g(x) = x^3 + 1$:

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

Next, we calculate $h(x) = (x^n - 1) : g(x)$, i.e. $h(x) = (x^6 - 1) : (x^3 + 1) = (x^6 + 1) : (x^3 + 1) = x^3 + 1$. Then the parity check matrix is as follows:

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

Cyclic codes

Example 4. Let C be a binary cyclic code of length 7 over F_2 with the generator polynomial: $g(x) = x^3 + x + 1$.

- a) Verify that the code C is cyclic.
- b) Find the generator matrix and parity check matrix for the given binary cyclic code C.

Hint for a): Note that every cyclic code is a polynomial code. Verify that g divides $x^7 - 1$.

Solution.

- a) We easily verify that $x^7 - 1 = 1 + x^7 = (1 + x + x^3)(1 + x + x^2 + x^4)$ over F_2, so $g(x)$ divides $x^7 + 1$ (or $x^7 - 1$) and thus the code C is a cyclic $[7, 4]$ code.
- b) We have $n = 7$. Note that we have defined k, such that $deg(g(x)) = n - k$, then $n - k = 3$ and hence $k = 4$. The generator matrix is obtained immediately from the knowledge of the coefficients of the polynomial $g(x) = x^3 + x + 1$:

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

Next, we calculate $h(x) = (x^n - 1) : g(x)$, i.e.,
$h(x) = (x^7 - 1) : (x^3 + x + 1) = 1 + x + x^2 + x^4$. Then the parity check matrix is as follows:

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

Cyclic codes

How to get all cyclic codes of a given length n?
All we need to do is to find all the factors of $x^n - 1$.

Note: Every cyclic code is a polynomial code. So we can use the following statement from the lecture:
A polynomial code is cyclic if and only if its generator polynomial divides $x^n - 1$, where n is the length of the code.

Cyclic codes

Example 5. Find all binary cyclic codes of length 3 over F_2.

Solution. If we want to determine $g(x)$, then we need to find the factorization of the polynomial $x^3 - 1$ over the field F_2. Note that $x^3 - 1 = (x+1)(x^2+x+1)$. Let $R_3 = F_2[x]/(x^3-1)$. We get the following results:

generator polynomial	code in R_3
1	R_3
$x+1$	$\{0, 1+x, x+x^2, 1+x^2\}$
x^2+x+1	$\{0, 1+x+x^2\}$
x^3-1	$\{0\}$

Finite fields

Definition

Let us have a finite field F_q and non-zero element $a \in F_q$. The smallest natural number n such that $a^n = 1$, is called the order of the element.

Consider the field F_{2^3}. This field is formed by polynomials over F_2 modulo the irreducible polynomial $x^3 + x + 1$. It contains the elements $\left\{0, 1, x, x+1, x^2, x^2+1, x^2+x, x^2+x+1\right\}$. The characteristic of this field is $p = 2$. All elements except 0 and 1 have order $n = q - 1 = 8 - 1 = 7$, and hence they are all primitive.

When we work with Reed-Solomon codes, it will be convenient for us to represent the non-zero elements of the finite field as powers of the primitive element (i.e., the generator of F_q^*). Let's choose one of the primitive elements in the field F_{2^3} (for example x) and denote it by α. By the element α^2 we mean the product $\alpha \cdot \alpha = x \cdot x = x^2$. We continue further with $\alpha^3 = \alpha^2 \cdot \alpha = x^2 \cdot x = x + 1$. We list the remaining powers in the following table:

α	x
α^2	x^2
α^3	$x+1$
α^4	x^2+x
α^5	x^2+x+1
α^6	x^2+1
α^7	1

Example 6. Decide whether there is a Reed-Solomon code with parameters $[7,5,3]_q$. If such a code exists, find its parity check matrix.

Solution.
We are looking for q, for which $7 = n = q - 1$, apparently it is exactly $q = 2^3$. Let us represent the elements of the field F_8 using the root α of the polynomial $x^3 + x + 1$ irreducible over F_2, so $F_8 = \left\{a_0 + a_1\alpha + a_2\alpha^2 \mid a_i \in F_2\right\}$. Since the group F_8^* is cyclic, every non-unit element is of order 7, therefore let's calculate the matrix

$$H = \begin{pmatrix} 1 & \alpha & \alpha^2 & \alpha+1 & \alpha^2+\alpha & \alpha^2+\alpha+1 & \alpha^2+1 \\ 1 & \alpha^2 & \alpha^2+\alpha & \alpha^2+1 & \alpha & \alpha+1 & \alpha^2+\alpha+1 \end{pmatrix}$$

Reed-Solomon codes

Example 7. Consider the finite field F_5 and let $\alpha = 2$. Find:

- generator polynomial for $RS(4,2)$ (i.e., length is $n = 4$ and dimension is $k = 2$)
- generator matrix for $RS(4,2)$
- check parity matrix for $RS(4,2)$.

Solution.

- Consider a finite field F_5 and $\alpha = 2$. It is easy to check that $ord(\alpha) = 4$, and α is therefore a primitive element for F_5^*.
  Note: we create the generator polynomial $g(x)$ of the RS code using the following formula:

$$g(x) = (x - \alpha)(x - \alpha^2)\dots(x - \alpha^{n-k}),$$

  where α is a primitive element.
  Then the generator polynomial is:

$$g(x) = (x - 2)(x - 4) = 3 + 4x + x^2.$$

- We can also write the generator matrix for $RS(4, 2)$:

$$G = \begin{pmatrix} 3 & 4 & 1 & 0 \\ 0 & 3 & 4 & 1 \end{pmatrix}$$

Reed-Solomon codes

- We know the generator matrix and we need to find the parity check matrix for $RS(4,2)$. First, we modify the generator matrix into standard form and obtain the following matrix:

$$\begin{pmatrix} 1 & 0 & 3 & 4 \\ 0 & 1 & 3 & 2 \end{pmatrix}$$

  Now we have a generator matrix of the form $G = (I \mid A)$, then its parity check matrix is $H = (-A^T \mid I)$, where I is the identity matrix. In our case

$$A = \begin{pmatrix} 3 & 4 \\ 3 & 2 \end{pmatrix}$$

Then we get the following parity check matrix

$$H = (-A^T \mid I) = \begin{pmatrix} 2 & 2 & 1 & 0 \\ 1 & 3 & 0 & 1 \end{pmatrix}$$

Binary Reed-Muller codes

The generator matrix of the Reed-Muller code of order r creates a code of length 2^m. The generator matrix of the Reed-Muller code can defined as a matrix consisting of $r+1$ partial submatrices:

$$G = \begin{pmatrix} G_0 \\ G_1 \\ \vdots \\ G_r \end{pmatrix},$$

where G_0 is a vector of length $n = 2^m$, which contains only ones (i.e., (1,...,1)), matrix G_1 has dimension $m \times 2^m$ and its columns are binary representations of the numbers $0, 1, \ldots, n$, where the leftmost column is $(0, \ldots, 0)^T$, and the rightmost column is $(1, \ldots, 1)^T$. The other submatrices G_l then have the size of $\binom{m}{l}$ rows and 2^m columns, with the fact that its rows are made up of arbitrary but different products (the $*_{\mod 2}$ operation applied component by component) of l rows of matrix G_1.

Example 8. Construct the generator matrix of the binary Reed-Muller code $R(2,3)$ and determine its length n.

Solution. We have $r = 2$ and $m = 3$, then $n = 2^m = 8$. First, we find out how many submatrices the generator matrix should consist of. For $r = 2$, the matrix G will be composed of $r + 1$ submatrices, that is

$$G = \begin{pmatrix} G_0 \\ G_1 \\ G_2 \end{pmatrix},$$

Binary Reed-Muller codes

Since the length of the code is $n = 8$, the generator matrix must have 8 columns. The construction of the matrix G_0 is trivial. Indeed, this matrix is only a single row vector with eight single zeros. Creating the matrix G_1 is also simple.The numbers 0 to 7 are written in binary form and are put into the matrix column by column. We get

$$G_0 = (11111111)$$

$$G_1 = \begin{pmatrix} 00001111 \\ 00110011 \\ 01010101 \end{pmatrix}.$$

The last step is to construct the submatrix G_2. Its rows are always formed by the product of any two rows of the matrix G_1, with the fact that no combination of multiplied rows may be repeated. We then determine the number of rows by calculating the expression $\binom{3}{2} = 3$. The products of rows 1 and 2 were selected for the first row, 2 and 3 for the second row, and 3 and 1 for the third row of the matrix G_2.

$$G_2 = \begin{pmatrix} 00000011 \\ 00010001 \\ 00000101 \end{pmatrix}.$$

Binary Reed-Muller codes

The resulting matrix G then has the form after the composition of the submatrices G_0, G_1, G_2

$$G = \begin{pmatrix} 11111111 \\ 00001111 \\ 00110011 \\ 01010101 \\ 00000011 \\ 00010001 \\ 00000101 \end{pmatrix}.$$

Example 9. Consider the Reed-Muller code $R(2,4)$. Find the generator matrix of the code $R(2,4)$.

Solution 1. We have $r = 2$ and $m = 4$, then $n = 2^m = 16$. First, we find out how many submatrices the generator matrix should consist of. For $r = 2$, the matrix G will be composed of $r + 1$ submatrices, that is

$$G = \begin{pmatrix} G_0 \\ G_1 \\ G_2 \end{pmatrix},$$

Binary Reed-Muller codes

Since the length of the code is $n = 16$, then the generator matrix must have 16 columns. The construction of the matrix G_0 is trivial. This matrix is only a one row vector with 16 one zeros. Creating the matrix G_1 is also simple. The numbers 0 to 15 are written in binary form and are put into the matrix column by column. We get

$$G_0 = (1111111111111111)$$

$$G_1 = \begin{pmatrix} 0000000011111111 \\ 0000111100001111 \\ 0011001100110011 \\ 0101010101010101 \end{pmatrix}.$$

Binary Reed-Muller codes

The last step is to construct the submatrix G_2. Its rows are always formed by the product of any two rows of the matrix G_1, with the fact that no combination of multiplied rows may be repeated. We then determine the number of rows by calculating the expression $\binom{4}{2} = 6$. The products of rows 1 and 2 were chosen for the first row, 3 and 1 for the second and 2 and 3 for the third row, 1 and 4 for the fourth row, 2 and 4 to the fifth row and 3 and 4 to the sixth row of the matrix G_2.

$$G_2 = \begin{pmatrix} 0000000000001111 \\ 0000000000110011 \\ 0000001100000011 \\ 0000000001010101 \\ 0000010100000101 \\ 0001000100010001 \end{pmatrix}.$$

Binary Reed-Muller codes

The resulting matrix G then has the form after the composition of the submatrices G_0, G_1, G_2

$$G = \begin{pmatrix} 1111111111111111 \\ 0000000011111111 \\ 0000111100001111 \\ 0011001100110011 \\ 0101010101010101 \\ 0000000000001111 \\ 0000000000110011 \\ 0000001100000011 \\ 0000000001010101 \\ 0000010100000101 \\ 0001000100010001 \end{pmatrix}$$

Solution 2. We have $r = 2$ and $m = 4$, then $n = 16$. Monomials in $F_2[x_1, x_2, x_3, x_4]$ of degree at most 2 are $\{1, x_1, x_2, x_3, x_4, x_1x_2, x_1x_3, x_1x_4, x_2x_3, x_2x_4, x_3x_4\}$. Vectors in F_2^{16} associated with these monomials are:

$$
\begin{aligned}
1 &\rightarrow & (11111111 \quad 11111111)\\
x_1 &\rightarrow & (01010101 \quad 01010101)\\
x_2 &\rightarrow & (00110011 \quad 00110011)\\
x_3 &\rightarrow & (00001111 \quad 00001111)\\
x_4 &\rightarrow & (00000000 \quad 11111111)\\
x_1x_2 &\rightarrow & (00010001 \quad 00010001)\\
x_1x_3 &\rightarrow & (00000101 \quad 00000101)\\
x_1x_4 &\rightarrow & (00000000 \quad 01010101)\\
x_2x_3 &\rightarrow & (00000011 \quad 00000011)\\
x_2x_4 &\rightarrow & (00000000 \quad 00110011)\\
x_3x_4 &\rightarrow & (00000000 \quad 00001111)
\end{aligned}
$$

Binary Reed-Muller codes

Therefore, the generator matrix of the $R(2,4)$ code is as follows:

$$\begin{pmatrix}
11111111 & 11111111 \\
01010101 & 01010101 \\
00110011 & 00110011 \\
00001111 & 00001111 \\
00000000 & 11111111 \\
00010001 & 00010001 \\
00000101 & 00000101 \\
00000000 & 01010101 \\
00000011 & 00000011 \\
00000000 & 00110011 \\
00000000 & 00001111
\end{pmatrix}$$

Linear codes

Example 10. Consider the following binary code
$C = \{(0,0,0),(1,0,1),(0,1,1),(1,1,0)\}$.

- Prove that C is a linear code.
- Find the distance d of the code C.
- Find the generator matrix G of the code C.

Solution:

- The vector $(0,0,0) \in C$, the addition operation of vectors from F_2^3 is closed and each element(vector) of C has an opposite element.
- We successively calculate the Hamming weight of all non-zero codewords and find that the minimum weight is equal to 2. According to the theorem (Let C be a linear code over F_q^n. Then $d(C) = wt(C)$), it follows that the minimum distance of the C code is equal to 2.

- The code size is 4, so $k = 2$, and the generator matrix G must have two rows. We can take, for example, the first two non-zero vectors and get:

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

Example 11. Consider the generator matrix G over the field F_3. Find the parity check matrix H of the linear code generated by the following matrix

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

Solution: We have the matrix

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

and then we subtract the first row from the 3rd row and get the following matrix:

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 2 & 2 & 2 \end{pmatrix}$$

Linear codes

Next, we multiply the 3rd row by 2 and subtract the second row from the first row. Then, we subtract the 3rd row from the 2nd row and get the following matrix:

$$G' = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix}$$

We use the following relation $H = (-A^T \mid I)$ and get

$$H = (-A^T \mid I_2) = \begin{pmatrix} 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 2 & 0 & 1 \end{pmatrix}$$

Cyclic codes

Example 12. Find all cyclic codes of length 4 over F_3.

Note: Every cyclic code is a polynomial code. So we can use the following statement from the lecture:
A polynomial code is cyclic if and only if its generator polynomial divides $x^n - 1$, where n is the length of the code.

Solution. If we want to determine $g(x)$, then we need to find the decomposition of the polynomial $x^4 - 1$ over the field F_3. Note that $x^4 - 1 = (x-1)(x+1)(x^2+1)$. We get the following results:

- Code (4,3) is generated by $x - 1 = x + 2$.
- Code (4,3) is generated by $x + 1$.
- Code (4,2) is generated by $x^2 + 1$.
- Code (4,2) is generated by $x^2 - 1 = x^2 + 2$.

- Code (4,1) is generated by $(x-1)(x^2+1) = x^3+2x^2+x+2$.
- Code (4,1) is generated by $(x+1)(x^2+1) = x^3+x^2+x+1$.

Cyclic codes

Example 13.
Find the generator matrix for the cyclic code (7,3) over F_2, whose generator polynomial is as follows: $x^4 + x^2 + x + 1$.

Solution. The generator matrix is obtained immediately from the knowledge of the coefficients of the polynomial $x^4 + x^2 + x + 1$:

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

Example 14. Find the generator matrix for the binary cyclic code of length 9 over F_2 with the generator polynomial:
$g(x) = x^6 + x^3 + 1.$

Solution. We have $n = 9$. Note that we have defined k, such that $deg(g(x)) = n - k$, then $n - k = 6$ and hence $k = 3$. The generator matrix is obtained immediately from the knowledge of the coefficients of the polynomial $g(x) = x^6 + x^3 + 1$:

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