dimension of code as vector space

A linear $[n, m, d]$-code over the finite field $\mathbb{F}_q$ is an m-dimensional subspace of $\mathbb{F}_q^n$, with minimal distance d.

$$\forall x, y \in C : \forall a, b \in \mathbb{F}_q : ax + by \in C$$

number of code words

a linear $[n, m, d]$ code over $\mathbb{F}_q$ is a regular (n, q^m, d) code over $\mathbb{F}_q$

Weight of a code word → total number of non zero elements in the word

In a linear code C the minimal distance is also the least weight of the code words of C

bcz if $d = d(x, y)$ then $x - y$ will be $\neq 0$ at exactly those places where x and y differ

$$\forall x, y \in C : d(x, y) = w(x - y)$$

also $(x - y) \in C$ so there will be a vector which has weight equal to the minimal distance.

Vice versa we have that $d(0, x) = w(x)$ so min dist = min weight.

//ripasso algebra lineare basi, generatori, dipendenza bla bla bla

Generator matrix:

m-dimensional code C with a base. We can make a $m \times n$ matrix of which the rows are the base vectors of the base.

We know there are q^m code words. For encoding we have to assign a codeword to a message.

We can do it by considering the message as a row matrix and multiplying it by the generator matrix.

Different bases different encodings.

Positional permutation is fine but symbolic breaks linearity

Two linear codes are equivalent if they can be obtained from each other using only positional permutations.

if $G = (\mathrm{Id}_m X)$ it's in normal position eg: $\begin{pmatrix} \overbrace{1\ 0}^{\mathrm{Id}_m} & 1 & 0 & 1 \\ 0\ 1 & 0 & 1 & 1 \end{pmatrix}$

If a code has such generator matrix it is called systematic code

this kind of codes are easier to decode

message m → codeword $mG = m[\mathrm{Id} X] = [m\ mX]$

the first m digits are equal to the message itself

theorem

every linear code is equivalent to a systematic code

proof: take any generator matrix A for the code C and run the algorithm:

A matrix mxn

for i=0 to m

swap rows so that the leftmost non zero element starting from row i is in row i

divide row i by this element

make zeroes in this column (where that element of row i is)

we obtain a mxn matrix A' reduced in row echelon form (ridotta a scala)

We know that A'=BA for some invertible matrix B bcz we only performed operation that do not nullify the determinant. So A' is also a generator matrix of C

To make C a systematic code:

p_i: position of the first non-zero element of A' on row i.

By construction the column p_i only has 1 one on row i, all others zero.

Now permute columns so that column p_i becomes row i

We obtain A'' which is a generator matrix in normal position.

The parity check matrix

We can also define the scalar product $\cdot$.

This maps every pair of code words onto $\mathbb{F}_q$ in a bilinear way

$\forall x := x_1 \cdots x_n,\ y = y_1 \cdots y_n \in \mathbb{F}_q^n : x \cdot y = x_1 y_1 \cdots x_n y_n \in \mathbb{F}_q$

if $x \cdot y = 0$ x and y are orthogonal

$\forall x, y, z \in \mathbb{F}_q^n : (x+y) \cdot z = x \cdot z + y \cdot z$

The orthogonal complement of a subset $S \subset \mathbb{F}_q^n$ is:

$$S^\perp := \{ x \in \mathbb{F}_q^n \mid \forall y \in S : x \cdot y = 0 \}$$

example: $C := \{00000, 11110, 01011, 10101\}$

$C^\perp := \{0000, 11110, 01010, 10100, 11011, 00111, 10011, 01101\}$

$C^\perp$ is still a linear code because of distributivity of the scalar product

$$x \cdot y = x \cdot z = 0 \Longrightarrow x(\lambda y + \mu z) = 0$$

this code is called orthogonal code or dual code

if C is a $[n, m, d]$ code $C^\perp$ is a $[n, m-n, e]$ code (e to be determined)

We consider G the generator matrix of C and H the one of $C^\perp$ then we know:

$$GH^T = 0 \wedge HG^T = 0$$

We call H the parity check matrix of C

If G is in standard form $[I_m X]$ we can also choose a special form of the parity check matrix

$H^T = \begin{pmatrix} -X \\ I_{m-n} \end{pmatrix}$ because $(I_m X)\begin{pmatrix} -X \\ I_{n-m} \end{pmatrix} = (X - X) = 0$

The parity check matrix is useful for decoding the code.

Because H is a generator matrix of the orthogonal code we have: $c \in C \iff Hc^T = 0$

for a random word $x \in \mathbb{F}_q^n$ we define its syndrome as

$$s(x) := Hx^t \in \mathbb{F}_q^{n-m}$$ if $s(x) = 0$ x is a codeword.

suppose we recieve r, assume it is in the form $c+e$, where c is the transmitted code word and e is the error vector. To remove the errors from r we have to find e

The syndrome of r only depends on e and not on c. So we should concentrate on the syndromes instead of the recieved words.

We should make a list of all possible syndromes and associate to each one the error vector with the least weight producing that syndrome (the most likely)

There is not always a unique error vector, so we also note the others in the list.

Algorithm: Encoding and decoding a linear code

C systematic $[n,m,d]$-code, G is a generator matrix in normal position with H the corresponding parity check matrix.

Encoding:

- take message $m \in \mathbb{F}_2^m$, send mG

Decoding:

- make list of syndromes and corresponding error vectors
- compute syndrome of recieved word $s := Hr^t$
- look at which error code s corresponds to.
  - If there is only one code subtract it from r.
    m is most likely the first m bits of $r-e$
  - if there are more codes they are equally probable so it's better to ask for retransmission.

---

example:

$[5,2,3]$-code $G := \begin{pmatrix} 1 0 1 1 0 \\ 0 1 0 1 1 \end{pmatrix}$

possible messages and their code words are then

message	code word
00	00000
10	10110
01	01011
11	11101

parity check matrix: $H := \begin{pmatrix} 1 0 1 0 0 \\ 1 1 0 1 0 \\ 0 1 0 0 1 \end{pmatrix}$

list of error vectors:

Syndrome	error vectors
000	00000
100	00100
010	00010
110	10000
001	00001
101	11000, 00101
011	01000
111	10001, 01100

suppose we recieve 11011, computing ssndrome gives $s: 110$, so $e := 10000$

correct code word then is 01011, message is 01.