

# On the distance distribution of systematic non-linear codes

Eleonora Guerrini

Department of Mathematics  
University of Trento  
Italy  
guerrini@science.unitn.it

Emmanuela Orsini

Department of Mathematics  
University of Milan  
Italy  
orsini@posso.dm.unipi.it

## Abstract

The most important families of non-linear codes are systematic. We provide a Gröbner bases technique to compute the distance distribution of systematic non-linear codes.

## Preliminaries

We recall some basic facts about polynomial rings and non-linear codes that will play a central role throughout this paper.

## Some polynomial rings

Let  $m \geq 1$  be a natural number. Let  $K$  be a field, let  $\bar{K}$  be the algebraic closure of  $K$  and let  $I$  be an ideal in the polynomial ring  $K[Y] = K[y_1, \dots, y_m]$ . We denote by  $E_q[Y]$  the following set of polynomial in  $K[Y]$ :

$$E_q[Y] = \{y_1^q - y_1, \dots, y_m^q - y_m\}$$

**Definition 1.** 1. Given a polynomial  $f \in K[Y]$ , we denote by  $\mathcal{V}(f)$  the set of all zeros of  $f$  in  $\bar{K}^m$ , i.e.

$$\mathcal{V}(f) = \{(a_1, \dots, a_m) \in \bar{K}^m \mid f(a_1, \dots, a_m) = 0\}.$$

2. Given an ideal  $I \subseteq K[Y]$  we denote by  $\mathcal{V}(I)$  the set of all zeros of  $I$ , i.e.

$$\mathcal{V}(I) = \{(a_1, \dots, a_m) \in \bar{K}^m \mid f(a_1, \dots, a_m) = 0 \text{ for all } f \in I\}$$

**Definition 2.** Let  $S \subseteq \bar{K}^m$ . Then the set of all polynomials  $f \in K[Y]$  such that  $f(a_1, \dots, a_m) = 0$  for any points  $(a_1, \dots, a_m) \in S$  forms an ideal in the polynomial ring  $K[Y]$ . This ideal is the **vanishing ideal** of  $S$  and is denoted by  $\mathcal{I}(S)$ .

Let  $\mathbb{F}_q$  be the finite field with  $q$  element and  $(\mathbb{F}_q)^m$  be the natural  $m$ -dimensional vector space over  $\mathbb{F}_q$ . Let  $1 \leq s \leq m-1$ . We fix in  $\mathbb{F}_q[y_1, \dots, y_s, t_1, \dots, t_{m-s}] = \mathbb{F}_q[Y, T]$ , the lexicographic order  $<$  with  $y_1 < y_2 < \dots < y_s < t_1 < \dots < t_{m-s}$ . Let  $I$  be an ideal in  $\mathbb{F}_q[Y, T]$  and  $G(I) \subset \mathbb{F}_q[Y, T]$  the minimal reduced Gröbner basis of  $I$ , w.r.t.  $<$  ordering, then we will use the following notation:

$$\mathcal{P}_{y_i} = \mathbb{F}_q[y_1, \dots, y_i] \setminus \mathbb{F}_q[y_1, \dots, y_{i-1}], \quad I_{y_i} = I \cap \mathcal{P}_{y_i}, \quad G_{y_i} = G \cap \mathcal{P}_{y_i}$$

$$\mathcal{P}_Y = \mathbb{F}_q[Y], \quad I_Y = I \cap \mathcal{P}_Y, \quad G_Y = G \cap \mathcal{P}_Y,$$

$$\mathcal{P}_{YT} = \mathbb{F}_q[Y, T] \setminus \mathbb{F}_q[Y], \quad I_{YT} = I \cap \mathcal{P}_{YT}, \quad G_{YT} = G \cap \mathcal{P}_{YT},$$

$$\mathcal{P}_{t_i} = \mathbb{F}_q[y_1, \dots, y_s, t_1, \dots, t_i] \setminus \mathbb{F}_q[y_1, \dots, y_s, t_1, \dots, t_{i-1}],$$

$$G_{t_i} = G \cap \mathcal{P}_{t_i}, \quad I_{t_i} = I \cap \mathcal{P}_{t_i}.$$

**Definition 3.** Let  $1 \leq f \leq m$  and  $p \in \mathbb{F}_q[y_1, y_2, \dots, y_m]$ . We say that  $p$  is a **simple  $t$ -monomial** if:  $p = y_{h_1} \dots y_{h_f}$  where  $h_1, \dots, h_f \in \{1, \dots, m\}$  and  $h_i \neq h_j, \forall i \neq j$ , i.e. a monomial in  $\mathbb{F}_q[y_1, \dots, y_m]$  such that  $\deg_{y_{h_i}} = 1, \forall 1 \leq i \leq f$ . We call  $\mathcal{M}_{m,f,q}$  the set of all simple  $f$ -monomials in  $\mathbb{F}_q[y_1, y_2, \dots, y_m]$ .

From now on we keep the reference to  $q$  implicit in our notation, so we will use  $\mathcal{M}_{m,t}$  instead of  $\mathcal{M}_{m,t,q}$ .

## Non-linear codes

**Definition 4.** Let  $k, n \in \mathbb{N}$  such that  $1 \leq k \leq n$ ,  $\phi: (\mathbb{F}_q)^k \rightarrow (\mathbb{F}_q)^n$  be an injective function and let  $C = \text{Im}(\phi)$ . We say that  $C$  is an  $(n, k, q)$  **code**. Any  $c \in C$  is called a **word** of the code.

**Definition 5.** Let  $\pi$  be the projection  $\pi: (\mathbb{F}_q)^n \rightarrow (\mathbb{F}_q)^k$  such that  $\pi(a_1, \dots, a_n) = (a_1, \dots, a_k)$ . We say that  $C$  is **systematic** if:  $(\pi \circ \phi)(v) = v$ , for any  $v \in (\mathbb{F}_q)^k$ .

We denote by  $\mathcal{C}(n, k, q)$  the class of all systematic  $(n, k, q)$  code.

## Weights in $(\mathbb{F}_q)^n$ and Gröbner bases

In this section we will show some relations between particular varieties in  $(\mathbb{F}_q)^n$  and Gröbner bases. Let  $\mathbb{F}_q[y_1, \dots, y_m]$  be a polynomial ring, we denote by  $\sigma_i$  the  $i$ -th elementary symmetric function of  $y_1, \dots, y_m$ . We also denote by  $I_{m,t}$  the following ideal:

$$I_{m,t} = \langle \sigma_t, \dots, \sigma_m, E_q[Y] \rangle \subset \mathbb{F}_q[y_1, \dots, y_m]$$

**Theorem 6.** Let  $m$  be an integer positive number such that  $m \geq 1$ . In  $\mathbb{F}_q[y_1, y_2, \dots, y_m]$  let  $t$  be an integer number such that  $1 \leq t \leq m$ . Let  $\sigma_i$  be the  $i$ -th symmetric function in  $y_1, y_2, \dots, y_m$  and  $P_i = \{c \in (\mathbb{F}_q)^m \mid w(c) = i\}$ . Then:

$$\mathcal{I}(P_0 \cup P_1 \dots \cup P_t) = \langle \sigma_{t+1}, \dots, \sigma_m, E_q[Y] \rangle = I_{m,t+1}$$

## A distance computing algorithm

### Gröbner basis of non-linear systematic codes

We apply the previous facts to give a structure for the Gröbner basis of the non-linear systematic codes. Let  $n, k \in \mathbb{N}$ , if  $C \in \mathcal{C}(n, k, q)$ , then we can view  $C$  as a set of points in  $(\mathbb{F}_q)^n \subset (\bar{\mathbb{F}}_q)^n$  and hence as a 0-dimensional variety, so that  $\mathcal{I}(C)$  is its vanishing ideal in  $\mathbb{F}_q[X, Z]$ . Our aim is to describe the reduced Gröbner basis of  $\mathcal{I}(C)$ .

**Theorem 7.** Let  $C \in \mathcal{C}(n, k, q)$  and  $G$  be the reduced Gröbner basis for  $\mathcal{I}(C)$ , w.r.t. the lexicographical order with  $x_1 < \dots < x_k < z_1 < \dots < z_{n-k}$ . Then  $G$  has the following structure:

$$G = \{E_q[X], z_1 + f_1, \dots, z_{n-k} + f_{n-k}\}$$

for some  $f_j \in \mathbb{F}_q[X], 1 \leq j \leq n-k$ . In particular

- $G_{x_i} = x_i^q - x_i, \quad 1 \leq i \leq k,$
- $G_{z_j} = z_j + f_j, \quad 1 \leq j \leq n-k.$

**Theorem 8.** Let  $\mathcal{A}_{k,n}$  be the set

$$\mathcal{A}_{k,n} = \{(f_1, \dots, f_{n-k}) \mid f_j: (\mathbb{F}_q)^k \mapsto (\mathbb{F}_q), 1 \leq j \leq n-k\}.$$

There is a bijection  $\mathcal{A}_{k,n} \leftrightarrow \mathcal{C}(n, k, q)$ , given by

$$(f_1, \dots, f_{n-k}) \leftrightarrow G = \{E_q[X], z_1 + f_1, \dots, z_{n-k} + f_{n-k}\}$$

### Weight distribution for non-linear systematic codes

**Definition 9.** Let  $C$  be a code in  $\mathcal{C}(n, k, q)$ . Let  $G(C) = \{E_q[X], f_1(X), \dots, f_{n-k}(X)\}$  the reduced Gröbner basis for  $C$ . Let  $t \in \mathbb{N}$  such that  $1 \leq t \leq n$ . We define the ideal  $\mathcal{W}_C^t \in \mathbb{F}_q[x_1, \dots, x_k]$  as follows:

$$\mathcal{W}_C^t = \{E_q[X], m(x_1, \dots, x_k, f_1(X), \dots, f_{n-k}(X)) \text{ for } m \in \mathcal{M}_{n,t}\}$$

A point in  $\mathcal{V}(\mathcal{W}_C^t)$  matches a codeword  $c$  in  $C$  with  $w(c) \leq t$ . We can easily derive the following

**Lemma 10.** Let  $C$  be in  $\mathcal{C}(n, k, q)$ . Let  $t \in \mathbb{N}$  such that  $1 \leq t \leq n$ . Then

$$A_{t-1} = |\mathcal{V}(\mathcal{W}_C^t)| \setminus |\mathcal{V}(\mathcal{W}_C^{t-1})|$$

### Distance of non-linear systematic codes

**Definition 11.** Let  $C \in \mathcal{C}(n, k, q)$  and  $f_1, \dots, f_{n-k}$ , as in Theorem 7. Then we denote by  $\mathcal{J}_C^t$ , with  $1 \leq t \leq n$ , the ideal in  $\mathbb{F}_q[x_1, x_2, \dots, x_k, \tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_k]$  generated by:

$$E_q[X] \cup \{m(x_i - \tilde{x}_i, \dots, f_j(x_1, \dots, x_k) - f_j(\tilde{x}_1, \dots, \tilde{x}_k), \dots) \mid 1 \leq i \leq k, 1 \leq j \leq n-k, m \in \mathcal{M}_{n,t}\}$$

**Definition 12.** In  $(\mathbb{F}_q)^k \times (\mathbb{F}_q)^k$  we denote by  $\mathcal{T}_k$  the **trivial variety**, i.e. the set of points  $a = (a_1, \dots, a_k, \tilde{a}_1, \dots, \tilde{a}_k)$  such that  $a_i = \tilde{a}_i, 1 \leq i \leq k$ .

**Lemma 13.** Let  $n, k \in \mathbb{N}$ . Then  $\forall C \in \mathcal{C}(n, k, q)$  and  $1 < t \leq n$ , we have

$$\mathcal{V}(\mathcal{J}_C^t) \neq \mathcal{T}_k \iff \exists c_1, c_2 \in C \text{ such that } d(c_1, c_2) \leq t-1$$

**Corollary 14.** Let  $n, k \in \mathbb{N}$ . Let  $q$  an integer positive prime number. Then  $\forall C \in \mathcal{C}(n, k, q)$  and  $1 < t \leq n$ ,

$$\mathcal{V}(\mathcal{J}_C^t) = \mathcal{T}_k \iff d(C) \geq t$$

From the Corollary 14, an algorithm is directly designed to compute the distance of any  $C \in \mathcal{C}(n, k, q)$ .

```

j = 1
While  $\mathcal{V}(\mathcal{J}_C^j) = \mathcal{T}_k$  do
  j := j + 1;
Output j

```

### The distance distribution for a code in $\mathcal{C}(n, k, q)$

Actually, Lemma 13 give a method to compute the distance distribution for a code in  $\mathcal{C}(n, k, q)$ . Indeed, let  $C \in \mathcal{C}(n, k, q)$  and  $1 < t \leq n$ , we showed that a point in  $\mathcal{V}(\mathcal{J}_C^t)$  is a pair of codewords with Hamming distance less than  $t$ . From that, we have the following:

**Corollary 15.** Let  $C \in \mathcal{C}(n, k, q)$ , let  $1 < t \leq n$ . Then

$$\mathcal{V}(\mathcal{J}_C^t) = \{(c_1, c_2) \mid s.t. c_1, c_2 \in C, d(c_1, c_2) \leq t\}$$

Let  $c_1, c_2 \in C$ , such that  $d(c_1, c_2) = i$ , for  $1 \leq i \leq t-1$ ; in  $\mathcal{J}_C^t$  there is both the points that match the pair of codewords  $(c_1, c_2)$  and the pair  $(c_2, c_1)$ . Moreover  $\mathcal{J}_C^t$  contains also the trivial variety. We can derive the following

**Lemma 16.** Let  $C \in \mathcal{C}(n, k, q)$ , let  $1 < t \leq n$ . Then

$$A_1 \cup A_2 \cup \dots \cup A_{t-1} = \frac{|\mathcal{V}(\mathcal{J}_C^t)| \setminus |\mathcal{T}_k|}{2}$$

**Example 17.** Let  $C = \{[0, 0, 0, 0], [0, 1, 0, 0], [0, 2, 0, 0], [1, 0, 0, 0], [1, 1, 0, 2], [1, 2, 0, 2], [2, 0, 0, 2], [2, 1, 0, 0], [2, 2, 0, 0]\}$  be an  $\mathcal{C}(4, 2, 3)$  code. The distance distribution of  $C$  is:

$$A_1 = 8, \quad A_2 = 20, \quad A_3 = 8, \quad A_4 = 0$$

We want to compute all the non ordered pairs of codewords  $(c_1, c_2)$  with  $d(c_1, c_2) \leq 2$ . To do this we calculate the ideal  $\mathcal{J}_C^3$ .

$$\mathcal{J}_C^3 = \langle x_2^2 \tilde{x}_1^2 \tilde{x}_2, x_1^2 x_2 \tilde{x}_1^2 \tilde{x}_2, \tilde{x}_2^3 - \tilde{x}_2, x_2^3 - x_2, \tilde{x}_1^3 - \tilde{x}_1, x_1^3 - x_1, x_2 \tilde{x}_1^2 \tilde{x}_2^2, x_1 \tilde{x}_1^2 \tilde{x}_2^2, x_1 x_2 \tilde{x}_1 \tilde{x}_2^2 \rangle$$

Using the Computer Algebra System Singular, we find that  $|\mathcal{V}(\mathcal{J}_C^3)| = 65$ . Since  $|\mathcal{T}_2| = 9$ , we have  $A_1 \cup A_2 = (65 - 9)/2 = 28$ .

From the previous results, we can easily derive a method to compute the distance distribution for a code in  $\mathcal{C}(n, k, q)$ .

**Lemma 18.** Let  $C \in \mathcal{C}(n, k, q)$ , let  $1 < t \leq n$ . Then

$$A_t = \frac{|\mathcal{V}(\mathcal{J}_C^{t+1})| \setminus |\mathcal{V}(\mathcal{J}_C^t)|}{2}$$

**Example 19.** Using the same code of 17 we want to find  $A_2$ :

$$\mathcal{J}_C^2 = \langle \tilde{x}_1 \tilde{x}_2 - x_1 \tilde{x}_2 - x_2 \tilde{x}_1 + x_1 x_2, \tilde{x}_2^3 - \tilde{x}_2, x_2^3 - x_2, \tilde{x}_1^3 - \tilde{x}_1, x_1^3 - x_1, x_1 \tilde{x}_2^2 - x_1 x_2^2, x_1 x_2^2 \tilde{x}_2 - x_1 \tilde{x}_2^2, \tilde{x}_1^2 - x_2^2 \tilde{x}_1 - \tilde{x}_1 + x_1 x_2^2 - x_1^2 + x_1, x_1 x_2^2 \tilde{x}_1 - x_1^2 \tilde{x}_1 + x_1 \tilde{x}_1 - x_1^2 \tilde{x}_2^2 - x_1^2 + x_1, x_1^2 x_2 \tilde{x}_1 + x_1 x_2 \tilde{x}_1 - x_1^2 x_2 - x_1 x_2 \rangle$$

As in the previous example we use the Computer Algebra System Singular to find that  $|\mathcal{V}(\mathcal{J}_C^2)| = 25$ . Applying Lemma 18, we obtain  $A_2 = (65 - 25)/2 = 20$ .

## References

- [1] Eleonora Guerrini, Emmanuela Orsini, Massimiliano Sala, Computing the distance distribution of systematic non-linear codes, Work in Progress, 2006.
- [2] Preparata, Franco, A class of optimum nonlinear double-error correcting codes, Inform. Control, vol.13, pp 378-400, 1968.