The non-gap sequence of a subcode of a generalized Reed-Solomon code

Irene Márquez-Corbella[1], Edgar Martínez-Moro[2], and Ruud Pellikaan[3]

[1] Department of Algebra, Geometry and Topology, University of Valladolid,
Prado de la Magdalena s/n, 47005 Valladolid, Spain.
imarquez@agt.uva.es

[2] Department of Applied Mathematics, University of Valladolid,
Campus Duques de Soria, E-42004 Soria, Spain.
edgar@maf.uva.es

[3] Department of Mathematics and Computing Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands.
g.r.pellikaan@tue.nl

Appeared in: Proceedings 7th International Workshop on Coding and Cryptography
April 11-15, Paris, pp. 183–193, 2011

Abstract. This paper addresses the question of how often the square code of an arbitrary l-dimensional subcode of the code $\mathrm{GRS}_k(\mathbf{a}, \mathbf{b})$ is exactly the code $\mathrm{GRS}_{2k-1}(\mathbf{a}, \mathbf{b} * \mathbf{b})$. To answer this question we first introduce the notion of gaps of a code which allows us to characterize such subcodes easily. This property was first stated and used in [10] where Wieschebrink applied the Sidelnikov-Shestakov attack [8] to brake the Berger-Loidreau cryptostystem [1].

Keywords: Berger-Loidreau cryptosystem, square code, GRS codes, gaps of a code.

1 Introduction

The notion of a *Public-Key Cryptosystem* (PKC) was first introduced in 1976 by Diffie and Hellman in [3]. Most PKC are based on hard number-theoretic problems such as integer factorization (like the RSA cryptosystem) or taking discrete logarithms in finite groups (like the El Gamal cryptosystem).

In [5] McEliece introduced the first PKC based on the theory of error-correcting codes which resist precisely the attacks to which the RSA and El Gamal cryptosystem are vulnerable. This property makes McEliece scheme an interesting candidate for post-quantum cryptography (see [7, 2] for an overview of the state of the art). Later, Niederreiter presents a dual version of the previous cryptosystem which is equivalent in terms of security (see [4]). In its original paper [6] Niederreiter proposed the class of Generalized Reed-Solomon (GRS) codes over $\mathbb{F}_{2^m}$. However, Sidelnikov and Shestakov in [8] introduced an algorithm that, in polynomial time, permits us to discover the structure of the secret

GRS code used in the cryptosystem. Therefore, the initial Niederreiter scheme is completely broken.

That is why Berger and Loidreau in [1] propose another version of the Niederreiter scheme which is designed to resist the Sidelnikov-Shestakov attack. The main idea of this variant is to work with subcodes of the original GRS code rather than using the complete GRS code.

However, in [9] Wieschebrink presents the first feasible attack to the Berger-Loidreau cryptosystem that allows us to recover the secret key if the chosen subcode is large enough, which is impractical for small subcodes. Furthermore, in [10] Wieschebrink notes that if the double code of a subcode of a GRS code of parameters $[n, k]$ is itself a GRS code of dimension $2k-1$ (what he says that seems to happen with high probability) then we can apply the Sidelnikov-Shestakov attack and thus reconstruct the secret key in polynomial time.

The main task of this paper is to confirm the previous question and give a characterization of the possible parameters that should be used to avoid attacks on the Berger-Loidreau cryptosystem.

The structure of this paper is as follows. First we give a brief review of basic concepts from Coding Theory that are relevant to this work and thus establish the notation that will be used throughout the paper. Furthermore, after discussing about the basic attributes and structure of GRS Codes we introduce the non-gap sequence associated to subcodes of such codes and we define some properties that characterize these subcodes in terms of this sequence. The final result of this section even allows us to count the number of subcodes of a GRS code if we identify all possible associated non-gap sequence.

The main objective of the second section is to study the probability that an arbitrary subcode of a GRS code is itself a GRS code. To achieve our goal it would be enough to analyze the non-gaps sequences associated to subcodes with the required properties, which provides an upper bound on the number of such subcodes and, consequently, we could estimate the probability of the occurrence that we are looking for.

Finally we introduce the square code of a subcode of a GRS code. Wieschebrink in [10] stated that this code is often a GRS code. In the last section we give some clues on how to solve this question but the final result will appear in a subsequent paper. However, we establish some properties that the non-gap sequence of the analyzed subcode must verify to have a square code which is a GRS code and has dimension $2k - 1$. Therefore, the weak subcodes for the Berger-Loidreau cryptosystem are completely characterized.

In order to shorten this abstract we will show all the results without proofs.

2 Gaps of a code

Let us start fixing the notation and introducing some basic definitions and some known results from coding theory. Let $\mathbb{F}_q$ be a finite field with q elements and let $n,\ k \in \mathbb{N}$ such that $1 \leq k \leq n \leq q$.

We define the set $L_k = \{f \in \mathbb{F}_q[X] : \deg(f(X)) \leq k-1\}$. Then for each $\mathbf{a},\ \mathbf{b} \in \mathbb{F}_q^n$ the evaluation map at these elements is given by:

$$\begin{array}{rcl} \mathrm{ev}_{\mathbf{a},\mathbf{b}} : L_k & \longrightarrow & \mathbb{F}_q^n \\ f & \mapsto & \mathrm{ev}_{\mathbf{a},\mathbf{b}}(f(X)) = (f(a_1)b_1, \ldots, f(a_n)b_n) \end{array}$$

We will denote the map $\mathrm{ev}_{\mathbf{a},\mathbf{b}}$ by $\mathrm{ev}_{\mathbf{a}}$ if $\mathbf{b}$ is the all one vector. For all vector $\mathbf{b} \in (\mathbb{F}_q \setminus \{0\})^n$ this evaluation map is injective, since $f \in L_k$ has at most $k \leq n$ zeros.

Let $\mathbf{a}$ be a vector of mutually distinct elements of $\mathbb{F}_q$ and $\mathbf{b}$ a vector consisting of n nonzero entries of $\mathbb{F}_q$ then the *generalized Reed-Solomon* code (or *GRS* code) is defined by $\mathrm{GRS}_k(\mathbf{a},\mathbf{b}) := \{\mathrm{ev}_{\mathbf{a},\mathbf{b}}(f(X)) : f \in L_k\}$.

That is, for every codeword $\mathbf{c} \in \mathrm{GRS}_k(\mathbf{a},\mathbf{b})$ there exists a unique polynomial $f_{\mathbf{c}} \in L_k$, known as *polynomial associated to* $\mathbf{c}$, such that $\mathbf{c} = \mathrm{ev}_{\mathbf{a},\mathbf{b}}(f_{\mathbf{c}}(X))$.

For $\mathbf{a}, \mathbf{b} \in \mathbb{F}_q^n$ we define the Schur or star product $\mathbf{a} * \mathbf{b} \in \mathbb{F}_q^n$ by:

$$\mathbf{a} * \mathbf{b} = (a_1 \cdot b_1, \ldots, a_n \cdot b_n) .$$

Let $\mathbf{1}$ be the all one vector. Then $\mathrm{GRS}_k(\mathbf{a},\mathbf{b}) = \mathbf{b} * \mathrm{GRS}_k(\mathbf{a},\mathbf{1})$. Furthermore

$$\mathrm{ev}_{\mathbf{a},\mathbf{b}}(f(X)g(X)) = \mathrm{ev}_{\mathbf{a},\mathbf{1}}(f(X)) * \mathrm{ev}_{\mathbf{a},\mathbf{b}}(g(X)).$$

From now on, let l be an integer such that $1 \leq l \leq k \leq n \leq q$ then $\mathcal{C}$ denotes an l-dimensional subcode of the code $\mathrm{GRS}_k(\mathbf{a},\mathbf{b})$ and we denote by $\mathcal{C}_i = \mathcal{C}_i(\mathbf{a},\mathbf{b}) := \mathcal{C} \cap \mathrm{GRS}_i(\mathbf{a},\mathbf{b})$.

We have the following nice embedding property:

$$\mathcal{C}_0 \subseteq \mathcal{C}_1 \subseteq \ldots \subseteq \mathcal{C}_k = \mathcal{C} \cap \mathrm{GRS}_k(\mathbf{a},\mathbf{b}) = \mathcal{C}.$$

Definition 1. *$i \in \mathbb{Z}_{\geq 0}$ is called an $(\mathbf{a},\mathbf{b})$-gap of the code $\mathcal{C}$ or simply a gap of $\mathcal{C}$ if $\mathcal{C}_i = \mathcal{C}_{i+1}$.*

The next Proposition show us how to identify all the $(\mathbf{a},\mathbf{b})$ non-gaps of any l-dimensional subcode of the code $\mathrm{GRS}_k(\mathbf{a},\mathbf{b})$.

Proposition 1. *$i \in \mathbb{Z}_{\geq 0}$ is an $(\mathbf{a},\mathbf{b})$ non-gap of $\mathcal{C}$ if and only if there exists $f \in \mathbb{F}_q[X]$ with $\deg(f(X)) = i$ such that $\mathrm{ev}_{\mathbf{a},\mathbf{b}}(f(X)) \in \mathcal{C}$.*

We define an associated $(\mathbf{a},\mathbf{b})$ non-gap sequence of $\mathcal{C}$ by

$$\mathcal{I}(\mathcal{C},\mathbf{a},\mathbf{b}) = \mathcal{I}(\mathcal{C}) = \{i \in \mathbb{Z}_{\geq 0} : i \text{ is a non-gap of } \mathcal{C}\}.$$

Using the previous characterization proposition we have the following result that allows us to define any subcode $\mathcal{C} \subseteq \mathrm{GRS}_k(\mathbf{a},\mathbf{b})$ as the subspace generated by polynomials in $\mathbb{F}_q[X]$ whose degree is an element of the associated $(\mathbf{a},\mathbf{b})$ non-gap sequence of $\mathcal{C}$.

Corollary 1. *Let $\mathcal{C}$ be an l-dimensional subcode of the code $\mathrm{GRS}_k(\mathbf{a},\mathbf{b})$ with associated non-gap sequence $\mathcal{I}(\mathcal{C})$. Then:*

1. $\mathcal{I}(\mathcal{C}) = \{i : \exists f \in \mathbb{F}_q[X] \text{ with } \deg(f(X)) = i < k \text{ such that } \mathrm{ev}_{\mathbf{a},\mathbf{b}}(f(X)) \in \mathcal{C}\}$.
2. $\mathcal{C} = \{\mathrm{ev}_{\mathbf{a},\mathbf{b}}(f(X)) : f = 0 \text{ or } f \in \mathbb{F}_q[X] \text{ and } \deg(f(X)) \in \mathcal{I}(\mathcal{C})\}$.

Furthermore applying elementary operations to the above result we obtain a basis of any l-dimensional subcode $\mathcal{C} \subseteq \mathrm{GRS}_k(\mathbf{a}, \mathbf{b})$ just studying the associated $(\mathbf{a}, \mathbf{b})$ non-gap sequence of $\mathcal{C}$.

Proposition 2. *There is a set $\mathcal{I}$ consisting of the strictly increasing sequence of non-negative integers $i_1, \dots, i_l$ and there are polynomials $f_1, \dots, f_l$ in the unique normal form*

$$f_j(X) = X^{i_j} + \sum_{\substack{s<i_j \\ s\notin\mathcal{I}}} f_{j,s}X^s \in \mathbb{F}_q[X], \text{ for all } j = 1, \dots, l,$$

such that the evaluation of these elements with respect to $(\mathbf{a}, \mathbf{b})$ form a basis of the code $\mathcal{C}$. Furthermore $\mathcal{I}(\mathcal{C}) = \mathcal{I}$ and $\dim(\mathcal{C}) = |\mathcal{I}(\mathcal{C})|$.

Therefore Proposition 2 allows us to count the number of l-dimensional subcodes $\mathcal{C}$ by identifying all possible associated non-gaps sequences (i.e. analyzing all subsets of the set of integers $\{0, \dots, k-1\}$).

Proposition 3. *Let $\mathcal{I}$ be a set consisting of the strictly increasing sequence of non-negative integers $i_1, \dots, i_l$. Let*

$$\begin{aligned} e(\mathcal{I}) &= i_1 l + (i_2 - i_1 - 1)(l-1) + \cdots + (i_l - i_{l-1} - 1) \\ &= \sum_{s=1}^{l} (i_s - i_{s-1} - 1)(l - s + 1) \end{aligned}$$

where $i_0 = -1$. Then the number of l-dimensional subcodes of the code $\mathrm{GRS}_k(\mathbf{a}, \mathbf{b})$ over $\mathbb{F}_q$ with a given non-gap sequence $\mathcal{I}$ is equal to $q^{e(\mathcal{I})}$.

The number $e(\mathcal{I})$ is minimal and equal to 0 for $\mathcal{I} = \{0, 1, \dots, l-1\}$ and it is maximal and equal to $l(k-l)$ for $\mathcal{I} = \{k-l, \dots, k-2, k-1\}$.

Accordingly to a well-known result and by Proposition 3 we have that the number of l-dimensional subcodes of the code $\mathrm{GRS}_k(\mathbf{a}, \mathbf{b})$ over $\mathbb{F}_q$ is equal to the Gaussian binomial:

$$\frac{(q^k-1)(q^k-q)\cdots(q^k-q^{l-1})}{(q^l-1)(q^l-q)\cdots(q^l-q^{l-1})} := \begin{bmatrix} k \\ l \end{bmatrix}_q = \sum_{\substack{\mathcal{I}\subseteq\{0,\dots,k-1\} \\ |\mathcal{I}|=l}} q^{e(\mathcal{I})}.$$

Hence this number is polynomial in q with non-negative integers as coefficients.

3 GRS subcodes of GRS codes

In this section we study the particular case of l-dimensional subcodes $\mathcal{C}$ of the code $\mathrm{GRS}_k(\mathbf{a}, \mathbf{b})$ that are themselves GRS codes.

The simplest case is when $\mathcal{C} = GRS_l(\mathbf{a}, \mathbf{b})$ with $2 \leq l \leq k$. The result presented bellow characterize uniquely such subcodes.

Proposition 4. *$\mathcal{C} = \mathrm{GRS}_l(\mathbf{a}, \mathbf{b})$ if and only if $\mathcal{I}(\mathcal{C}) = \{0, \dots, l-1\}$.*

Corollary 2. *There is exactly one l-dimensional subcode of the code $\mathrm{GRS}_k(\mathbf{a}, \mathbf{b})$ over $\mathbb{F}_q$ with non-gap sequence $\{0, \dots, l-1\}$, that is $\mathrm{GRS}_l(\mathbf{a}, \mathbf{b})$.*

Another special case is when $\mathcal{C} = \mathrm{GRS}_l(\mathbf{a}, \mathbf{a}^i * \mathbf{b})$ with $i + l < k$. The following results allow us to give an upper bound on the number of such subcodes.

Proposition 5. *Let i_l be the largest non-gap of $\mathcal{C}$ and let $\mathbf{c} = \mathrm{ev}_{\mathbf{a}}(f(X))$ for $f \in \mathbb{F}_q[X]$ of degree i. If $i + i_l < k$, then $\mathcal{I}(\mathbf{c} * \mathcal{C}) = i + \mathcal{I}(\mathcal{C})$.*

Let us define $\mathbf{a}^i$ by induction: $\mathbf{a}^0 = \mathbf{1}$, $\mathbf{a}^1 = \mathbf{a}$ and $\mathbf{a}^{i+1} = \mathbf{a} * \mathbf{a}^i$.

Corollary 3. *If $i + l \leq k$, then the non-gap sequence of the code $\mathrm{GRS}_l(\mathbf{a}, \mathbf{a}^i * \mathbf{b})$ is equal to $\{i, i+1, \dots, i+l-1\}$.*

Note that the converse is not true in general.

Corollary 4. *The number of l-dimensional subcodes of the code $\mathrm{GRS}_k(\mathbf{a}, \mathbf{b})$ over $\mathbb{F}_q$ with l consecutive non-gaps is equal to*

$$\sum_{l=1}^{k} \sum_{i=0}^{k-l} q^{il}.$$

Now we consider the general case i.e. we consider $\mathcal{C} = \mathrm{GRS}_l(\mathbf{c}, \mathbf{d})$ an l-dimensional subcode of the code $\mathrm{GRS}_k(\mathbf{a}, \mathbf{b})$. First we give the necessary conditions that $\mathcal{C}$ must verified to be a subcode of the code $\mathrm{GRS}_l(\mathbf{a}, \mathbf{b})$ and how is its associated non-gap sequence. Then, with an additional assumption, we show the converse (which are the necessary conditions that the associated non-gap sequence of $\mathcal{C}$ must verified to be a subcode of the code $\mathrm{GRS}_k(\mathbf{a}, \mathbf{b})$). With these two results we can give an upper bound on the number of such subcodes.

Proposition 6. *Let $l \geq 2$ and let $\mathbf{a}$ be a vector of n mutually distinct elements of $\mathbb{F}_q$ and $\mathbf{b}$ a vector consisting of n nonzero entries of $\mathbb{F}_q$. Let $g_0, h_1 \in \mathbb{F}_q[X]$. Let $d_0 = \deg(g_0(X))$ and $d_1 = d_0 + \deg(h_1(X))$. Suppose that $\mathrm{ev}_{\mathbf{a}}(h_1(X)) = \mathbf{c}$ is a vector of n mutually distinct elements of $\mathbb{F}_q$ and $\mathrm{ev}_{\mathbf{a},\mathbf{b}}(g_0(X)) = \mathbf{d}$ is a vector consisting of nonzero entries. If $d_0 < d_1$ and $d_0 + (l-1)(d_1 - d_0) < k$, then $\mathrm{GRS}_l(\mathbf{c}, \mathbf{d})$ is an l dimensional subcode of the code $\mathrm{GRS}_k(\mathbf{a}, \mathbf{b})$ with $(\mathbf{a}, \mathbf{b})$ non-gap sequence:*

$$d_0, \dots, d_0 + j(d_1 - d_0), \dots, d_0 + (l-1)(d_1 - d_0).$$

If $2k - 2 < n$ and $l \geq 2$, then there is a converse of the above Proposition given by the following result.

Proposition 7. *Let $d_0 < d_1$ be the first two elements of the $(\mathbf{a}, \mathbf{b})$ non-gap sequence of $\mathcal{C}$ and suppose that $\mathcal{C} = \mathrm{GRS}_l(\mathbf{c}, \mathbf{d})$ where $\mathbf{c}$ is a vector of n mutually distinct elements of $\mathbb{F}_q$ and $\mathbf{d}$ is a vector consisting of nonzero entries. Then there exist $g_0, h_1 \in \mathbb{F}_q[X]$ such that $d_0 = \deg(g_0(X))$, $d_1 = d_0 + \deg(h_1(X))$, $\mathrm{ev}_{\mathbf{a}}(h_1(X)) = \mathbf{c}$ and $\mathrm{ev}_{\mathbf{a},\mathbf{b}}(g_0(X)) = \mathbf{d}$.*

Corollary 5. *If $2k-2<n$ and $2\le l\le k$. Then the number of l-dimensional subcodes of the code* $\mathrm{GRS}_k(\mathbf{a},\mathbf{b})$ *over $\mathbb{F}_q$ that are a GRS code is at most q^{k-l+3}.*

With the previous assumptions i.e. let $2\le l\le k$ and $2k-2<n\le q$, then Corollary 5 and Proposition 3 imply that the probability that an arbitrary l-dimensional subcode of the code $\mathrm{GRS}_k(\mathbf{a},\mathbf{b})$ is a generalized Reed-Solomon code is at most

$$\frac{q^{k-l+3}}{\begin{bmatrix} k \\ l \end{bmatrix}_q} \le \frac{q^{k-l+3}}{q^{l(k-l)}} = q^{-(l-1)(k-l)+3}$$

This fraction tends to zero for $k\to\infty$ or $(k-l)\to\infty$.

4 The square of a code

Definition 2. *We define the square code of a $[n,k]$ linear code C over $\mathbb{F}_q$ and we denoted by $\mathcal{D}=\langle C*C\rangle$ as the code generated by the set $\{\mathbf{r}_i*\mathbf{r}_j : 1\le i\le j\le k\}$ where $\mathbf{r}_1,\dots,\mathbf{r}_k$ denotes the rows of a generator matrix of the code C.*

Now let $\mathcal{C}$ be an l-dimensional subcode of the code $\mathrm{GRS}_k(\mathbf{a},\mathbf{b})$, let $\mathbf{r}_1,\dots,\mathbf{r}_l$ be the rows of a generator matrix of $\mathcal{C}$ and let $f_1,\dots,f_l$ be the polynomials associated to those rows, then $\mathbf{r}_i*\mathbf{r}_j$ for all $i,\ j\in\{1,\dots,l\}$ has the form:

$$\mathbf{r}_i*\mathbf{r}_j=\left(b_1^2f_i(a_1)f_j(a_1),\dots,b_n^2f_i(a_n)f_j(a_n)\right)$$

where $\deg(f_i(X)f_j(X))=\deg(f_i(X))+\deg(f_j(X))\le 2k-2$.

That is, the code $\mathcal{D}=\langle\mathcal{C}*\mathcal{C}\rangle:=\langle\mathbf{r}_i*\mathbf{r}_j:1\le i\le j\le j\rangle$ is a subcode of the code $\mathrm{GRS}_{2k-1}(\mathbf{a},\mathbf{b}*\mathbf{b})$.

Similarly to what we did in the previous section we denote by

$$\mathcal{D}_i=\mathcal{D}_i(\mathbf{a},\mathbf{b}*\mathbf{b}):=\mathcal{D}\cap\mathrm{GRS}_i(\mathbf{a},\mathbf{b}*\mathbf{b}).$$

In this case $i\in\mathbb{Z}_{\ge0}$ is called an $(\mathbf{a},\mathbf{b}*\mathbf{b})$ gap of $\mathcal{D}$ if $\mathcal{D}_i=\mathcal{D}_{i+1}$. Define the following set of non-gaps

$$\mathcal{J}(\mathcal{D})=\{j\in\mathbb{Z}_{\ge0} : j \text{ is an } (\mathbf{a},\mathbf{b}*\mathbf{b}) \text{ non-gap of } \mathcal{D}\}.$$

From now on we assume that $k,\ l$ are integers such that $2\le l\le k\le n\le q$ and $2k-1\le n$.

Remark 1. i is an $(\mathbf{a},\mathbf{b}*\mathbf{b})$ non-gap of $\mathcal{D}$ if and only if there exists a polynomial $g\in\mathbb{F}_q[X]$ with $\deg(g(X))=i$ such that $\mathrm{ev}_{\mathbf{a},\mathbf{b}*\mathbf{b}}(g(X))\in\mathcal{D}$. This result is directly consequence of Proposition 1.

Next Proposition shows the relationship between the $(\mathbf{a},\mathbf{b})$ non-gap sequence associated to the subcode $\mathcal{C}$, i.e. $\mathcal{I}(\mathcal{C})$, and the $(\mathbf{a},\mathbf{b}*\mathbf{b})$ non-gap sequence associated to the square code $\mathcal{D}$, i.e. $\mathcal{J}(\mathcal{D})$.

Proposition 8. *$\mathcal{I}(\mathcal{C}) + \mathcal{I}(\mathcal{C}) = \{i + j : i,\ j \in I(\mathcal{C})\} \subseteq \mathcal{J}(\mathcal{D})$. Furthermore:*

1. *If 0 is an $(\mathbf{a}, \mathbf{b})$ non-gap of $\mathcal{C}$ then $\mathcal{I}(\mathcal{C}) \subseteq \mathcal{J}(\mathcal{D})$.*
2. *Otherwise let i_1, i_l be the first and last element, respectively of the $(\mathbf{a}, \mathbf{b})$ non-gap sequence of $\mathcal{C}$. Let $\mathbf{c} = \mathrm{ev}_{\mathbf{a},\mathbf{b}}(f(X)) \in \mathcal{C}$ for $f \in \mathbb{F}_q[X]$ of degree i_1. Then, if $i_1 + i_l < 2k - 1$, we have that*

$$\mathcal{I}(\mathbf{c} * \mathcal{C}) = i_1 + \mathcal{I}(\mathcal{C}) \subseteq \mathcal{J}(\mathcal{D}).$$

However, the previous equality does not hold in general.

Proposition 9. *$\dim(\mathcal{D}) \leq \min\{2k - 1, \binom{l+1}{2}\}$. Furthermore:*

1. *If $\mathcal{D} = \mathrm{GRS}_r(\mathbf{a}, \mathbf{b} * \mathbf{b})$ then the associated $(\mathbf{a}, \mathbf{b})$ non-gap sequence of the code $\mathcal{C}$ verifies that $\mathcal{I}(\mathcal{C}) \subseteq \{0, \ldots, \lfloor \frac{r-1}{2} \rfloor\}$.*
2. *If $\mathcal{D} = \mathrm{GRS}_r(\mathbf{a}, \mathbf{a}^i * \mathbf{b} * \mathbf{b})$ then the associated $(\mathbf{a}, \mathbf{b})$ non-gap sequence of the code $\mathcal{C}$ verifies that*

$$\mathcal{I}(\mathcal{C}) \subseteq \left\{ \left\lfloor \frac{i}{2} \right\rfloor, \ldots, \left\lfloor \frac{i + r - 1}{2} \right\rfloor \right\}.$$

3. *If $\mathcal{D} = \mathrm{GRS}_r(\mathbf{c}, \mathbf{d})$ then the associated $(\mathbf{a}, \mathbf{b})$ non-gap sequence of the code $\mathcal{C}$ verifies that*

$$\mathcal{I}(\mathcal{C}) \subseteq \left\{ \left\lfloor \frac{d_0}{2} \right\rfloor, \ldots, \left\lfloor \frac{d_0 + (r-1)(d_1 - d_0)}{2} \right\rfloor \right\}$$

where $d_0 < d_1$ and $d_0 + (r - 1)(d_1 - d_0) < 2k - 1$.

The aim of this section is to obtain a characterization of the l-dimensional subcodes $\mathcal{C}$ of the code $\mathrm{GRS}_k(\mathbf{a}, \mathbf{b})$ such that its square code is exactly the code $\mathrm{GRS}_{2k-1}(\mathbf{a}, \mathbf{b} * \mathbf{b})$.

The first possibility occurs when $\mathcal{C}$ is itself a GRS code. The following two results are related to this special case.

Proposition 10. *$\langle \mathrm{GRS}_k(\mathbf{a}, \mathbf{b}) * \mathrm{GRS}_l(\mathbf{a}, \mathbf{c}) \rangle = \mathrm{GRS}_{k+l-1}(\mathbf{a}, \mathbf{b} * \mathbf{c})$.*

In particular, $\mathrm{GRS}_{2l-1}(\mathbf{a}, \mathbf{b} * \mathbf{b})$ is the square of the code $\mathrm{GRS}_l(\mathbf{a}, \mathbf{b})$.

Corollary 6. *If the associated $(\mathbf{a}, \mathbf{b})$ non-gap sequence of the subcode $\mathcal{C}$ is precisely $\mathcal{I}(\mathcal{C}) = \{0, \ldots, l - 1\}$ then the $(\mathbf{a}, \mathbf{b} * \mathbf{b})$ non-gap sequence associated to the square code $\mathcal{D}$ is $\mathcal{J}(\mathcal{D}) = \{0, \ldots, 2l - 2\}$, that is $\mathcal{D} = \mathrm{GRS}_{2l-2}(\mathbf{a}, \mathbf{b} * \mathbf{b})$.*

The converse is not true in general.

Next Proposition determines a property that must verified all the elements of the $(\mathbf{a}, \mathbf{b})$ non-gap sequence of $\mathcal{C}$ to have that the square code $\mathcal{D}$ is exactly the code $\mathrm{GRS}_{2k-1}(\mathbf{a}, \mathbf{b} * \mathbf{b})$.

Proposition 11. *Let the increasing sequence $i_1, \ldots, i_l$ be an enumeration of the $(\mathbf{a}, \mathbf{b})$ non-gap sequence of $\mathcal{C}$. If the square of $\mathcal{C}$ is equal to $\mathrm{GRS}_{2k-1}(\mathbf{a}, \mathbf{b} * \mathbf{b})$, then*

$$|\{\ (u, v)\ :\ i_u + i_v \geq t \text{ and } 1 \leq u \leq v \leq l\ \}| \geq 2k - t - 1$$

for all $t = 0, \ldots, 2k - 2$.

Remark 2. If the square code of $\mathcal{C}$ is equal to $\mathrm{GRS}_{2k-1}(\mathbf{a}, \mathbf{b} * \mathbf{b})$, then:

1. The special case $t = 0$ of Proposition 11 implies that $2k - 1 \leq \binom{l+1}{2}$ which is in agreement with Proposition 10.
2. The cases $t = 2k - 2$, $t = 2k - 3$ and $t = 2k - 5$ imply that $i_l = k - 1$, $i_{l-1} = k - 2$ and $i_{l-2} \geq k - 4$.

Remark 3. If $\mathcal{I}(\mathcal{C}) = \{k - l, \ldots, k - 1\}$, then this non-gap sequence satisfies the conditions of Proposition 11 for all t. However not all the square codes of a subcode $\mathcal{C}$ of the code $\mathrm{GRS}_k(\mathbf{a}, \mathbf{b})$ with such associated non-gap sequence are exactly the code $\mathrm{GRS}_{2k-1}(\mathbf{a}, \mathbf{b} * \mathbf{b})$. A case example is the subcode

$$\mathcal{C} = \left\langle \mathrm{ev}_{\mathbf{a},\mathbf{b}}(X^i) \ : \ i = k - l, \ldots, k - 1 \right\rangle .$$

Assume that we have two polynomials $f(X), g(X) \in L_k$, that is, both polynomials can be written as:

$$f(X) = \textstyle\sum_{r=0}^{k-1} f_r X^r \ \text{ and } \ g(X) = \textstyle\sum_{s=0}^{k-1} g_s X^s$$

with $f_r, g_s \in \mathbb{F}_q$ for $r, s \in \{0, \ldots, k - 1\}$.

The product of these polynomials give us another polynomial in L_{2k-2}

$$f(X)g(X) = h(X) = h_0 + h_1 X + \ldots + h_{2k-2} X^{2k-2} \in L_{2k-2}.$$

This can be expressed in matrix form as follows:

$$R(f)S(g)^T = \begin{pmatrix} h_{2k-2} \\ h_{2k-3} \\ \vdots \\ h_{k-1} \\ \vdots \\ h_0 \end{pmatrix},$$

where $R(f)$ is a matrix of size $(2k - 1) \times (3k - 2)$ over $\mathbb{F}_q$ with the following form:

$$R(f) = \begin{pmatrix} f_0 & f_1 & \cdots & f_{k-1} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & f_0 & \cdots & f_{k-2} & f_{k-1} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & f_0 & f_1 & \cdots & f_{k-1} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & f_0 & f_1 & \cdots & f_{k-1} \end{pmatrix}$$

and $S(g)$ is a matrix of size $1 \times (3k - 2)$ over $\mathbb{F}_q$ with

$$S(g) = (\underbrace{0 \cdots 0}_{k-1} \ g_{k-1} \cdots g_0 \ \underbrace{0 \cdots 0}_{k-1}).$$

Now let $\mathcal{C}$ be an l-dimensional subcode of the code $\mathrm{GRS}_k(\mathbf{a},\mathbf{b})$ with associated $(\mathbf{a},\mathbf{b})$ non-gap sequence $\mathcal{I}(\mathcal{C})=\{i_1,\dots,i_l\}$ then by Proposition 2 there exists l polynomials $f_j\in L_k$ for $j\in\{1,\dots,l\}$ in normal form of degree i_j such that

$$\mathcal{C}=\langle \mathrm{ev}_{\mathbf{a},\mathbf{b}}(f_1(X)),\dots,\mathrm{ev}_{\mathbf{a},\mathbf{b}}(f_l(X))\rangle.$$

Furthermore we know that the elements $\mathrm{ev}_{\mathbf{a},\mathbf{b}*\mathbf{b}}(f_u(X)f_v(X))$ generate the double code of $\mathcal{C}$, denoted by $\mathcal{D}$ with $1\le u\le v\le l$.

If we denote by $g_{uv}(X)=g_{uv0}+g_{uv1}X+\dots+g_{uv(2k-2)}X^{2k-2}\in L_{2k-1}$ the polynomial obtained by multiplying the polynomials f_u and f_v for $1\le u\le v\le l$ then the following matrix is a generator matrix of the square code $\mathcal{D}$.

$$G_{\mathcal{D}}=\begin{pmatrix} g_{11(2k-2)} & \cdots & g_{1l(2k-2)} & g_{22(2k-2)} & \cdots & g_{2l(2k-2)} & \cdots & g_{ll(2k-2)} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ g_{110} & \cdots & g_{1l0} & g_{220} & \cdots & g_{2l0} & \cdots & g_{ll0} \end{pmatrix},$$

where $G_{\mathcal{D}}$ is a matrix of size $(2k-1)\times\binom{l+1}{2}$ over $\mathbb{F}_q$.

We define

$$R=(\underbrace{R(f_1),\dots,R(f_1)}_{l},\underbrace{R(f_2),\dots,R(f_2)}_{l-1},\dots,R(f_l))\in\mathbb{F}_q^{(2k-1)\times\binom{l+1}{2}(3k-2)},$$

$$S(f_1,\dots,f_l)=\begin{pmatrix} S(f_1) & 0 & 0\cdots & 0 \\ 0 & S(f_2) & 0\cdots & 0 \\ \vdots & \vdots & \vdots\ddots & \vdots \\ 0 & 0 & 0\cdots & S(f_l) \end{pmatrix}\in\mathbb{F}_q^{l\times l(3k-2)}.$$

and

$$S=\begin{pmatrix} S(f_1,\dots,f_l) & 0 & 0\cdots & 0 \\ 0 & S(f_2,\dots,f_l) & 0\cdots & 0 \\ \vdots & \vdots & \vdots\ddots & \vdots \\ 0 & 0 & 0\cdots & S(f_l) \end{pmatrix}\in\mathbb{F}_q^{\binom{l+1}{2}\times\binom{l+1}{2}(3k-2)}.$$

Then $RS^T=G_{\mathcal{D}}$.

Remark 4. The following properties are necessary conditions to have that the square code $\mathcal{D}$ of the code $\mathcal{C}$ is the code $\mathrm{GRS}_{2k-1}(\mathbf{a},\mathbf{b}*\mathbf{b})$.

1. $\mathcal{I}(\mathcal{C})=\{i_1,\dots,i_l\}\subseteq\{0,\dots,k-1\}$.
2. $i_l=k-1$, $i_{l-1}=k-2$ and $i_{l-2}\ge k-4$.
3. The matrix $G_{\mathcal{D}}$ has full rank, i.e. $\mathrm{rank}(R(f_1),\dots,R(f_l))=2k-1$.

We are aiming to prove in a subsequent paper that the square of almost all l-dimensional subcodes of $\mathrm{GRS}_k(\mathbf{a},\mathbf{b})$ is equal to $\mathrm{GRS}_{2k-1}(\mathbf{a},\mathbf{b})$.

References

1. T. Berger and P. Loidreau. *How to mask the structure of codes for a cryptographic use.* Designs, Codes and Cryptography, 35: 63–79, 2005.
2. D. J. Bernstein. *Grover vs. McEliece.* PQCrypto 2010, 36, 73–80, 2010.
3. W. Diffie and M. Hellman. *New Directions in Cryptography.* IEEE Transaction on Information Theory, IT-22, 644–654, 1976.
4. Y. X. Li, R. H. Deng and X. M. Wang. On the equivalence of McEliece's and Niederreiter's public-key cryptosystems. IEEE Transaction on Information Theory, IT-40, 27–273,1994.
5. R. J. McEliece. *A public-key cryptosystem based on algebraic coding theory.* DSN Progress Report, 42-44:114–116, 1978.
6. H. Niederreiter. *Knapsack-type crypto systems and algebraic coding theory.* Problems of Control and Information Theory, 15(2):159–166, 1986.
7. R. Overbeck and N. Sendrier. *Code-based cryptography.* Post-quantum cryptography, 6, 95–145, 2009.
8. V. M. Sidelnikov and S. O. Shestakov. *On the insecurity of cryptosystems based on generalized Reed-Solomon codes.* Discrete Math. Appl., 2:439–444, 1992.
9. C. Wieschebrink. *An attack on the modified Niederreiter encryption scheme.* In PKC 2006, Lecture Notes in Computer Science, volume 3958, 14–26, Berlin, 2006. Springer.
10. C. Wieschebrink. *Cryptoanalysis of the Niederreiter public key scheme based on GRS subcodes.* In Post-Quantum Cryptography, Lecture Notes in Computer Science, volume 6061, 6–72, Berlin, 2010. Springer.