A characterization of MDS codes that have an error correcting pair

I. Márquez-Corbella [1]   R. Pellikaan [2]

[1] Department of Algebra, Geometry and Topology, University of Valladolid.
Supported by a FPU grant AP2008-01598 by Spanish MEC.

[2] Department of Mathematics and Computing Science, Eindhoven University of Technology.

Code-based Cryptography Workshop (CBC) 2012

INTRODUCTION TO CODING THEORY

- An **$[n, k]$ linear code $\mathcal{C}$** over $\mathbb{F}_q$ is a k-dimensional subspace of $\mathbb{F}_q^n$.

  Its **size** is $M = q^k$, the **information rate** is $R = \frac{k}{n}$ and the **redundancy** is $n - k$.
- The **generator matrix** of $\mathcal{C}$ is a $k \times n$ matrix G whose rows form a basis of $\mathcal{C}$, i.e.
$$\mathcal{C} = \left\{ \mathbf{x}G \mid \mathbf{x} \in \mathbb{F}_q^k \right\}.$$
- The **parity-check matrix** of $\mathcal{C}$ is an $(n - k) \times n$ matrix H whose nullspace is generated by the codewords of $\mathcal{C}$, i.e.
$$\mathcal{C} = \left\{ \mathbf{y} \in \mathbb{F}_q^n \mid H\mathbf{y}^T = 0 \right\}.$$
- The **hamming distance** between $\mathbf{x}, \mathbf{y} \in \mathbb{F}_q^n$ is $d_H(\mathbf{x}, \mathbf{y}) = |\{i \mid x_i \neq y_i\}|$.
- The **minimum distance** of $\mathcal{C}$ is
$$d(\mathcal{C}) = \min \left\{ d_H(\mathbf{c}_1, \mathbf{c}_2) \mid \mathbf{c}_1, \mathbf{c}_2 \in \mathcal{C} \text{ and } \mathbf{c}_1 \neq \mathbf{c}_2 \right\}.$$

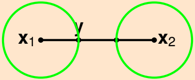


FIGURE: If $d(\mathcal{C}) = 3$

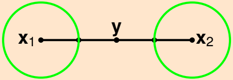


FIGURE: If $d(\mathcal{C}) = 4$

MDS CODES

One of the most fascinating chapters in all of coding theory

Let $\mathcal{C}$ be a linear code over $\mathbb{F}_q$, we will denote:

➜ Its length by $n(\mathcal{C})$ ➜ Its dimension by $k(\mathcal{C})$ ➜ Its minimum distance by $d(\mathcal{C})$

SINGLETON BOUND

$$d(\mathcal{C}) \leq n(\mathcal{C}) - k(\mathcal{C}) + 1$$

If the equality holds $\Longrightarrow$ $\mathcal{C}$ is an **MDS code**.

EXAMPLES

1. The **zero code** of length n (i.e. the $[n, 0, n+1]$ linear code) and **its dual** (i.e. $\mathbb{F}_q^n$ which has parameters $[n, n, 1]$).
2. The $[n, 1, n]$ **repetition code** over $\mathbb{F}_q$.
3. The **(Extended / Generalized) Reed-Solomon codes**.

F. J. MacWilliams, N. J. A. Sloane
The theory of error-correcting codes II.
North-Holland Mathematical Library, Vol 16.

MDS CODES

A collection of some properties characterizing MDS codes:

THEOREM: PROPERTIES OF MDS CODES

Let $\mathcal{C}$ be an $[n, k]$ code over $\mathbb{F}_q$. The following are equivalent:

1. $\mathcal{C}$ is MDS.
2. $\mathcal{C}^{\perp}$ is MDS.
3. Every k-tuple of columns of a generator matrix of $\mathcal{C}$ is independent.
4. Every set of k coordinates form an information set.
5. Every $n - k$-tuple of columns of a parity check matrix of $\mathcal{C}$ is independent.

Modifying Codes

- Let $\mathcal{C}$ be a linear $[n, k]$ code over $\mathbb{F}_q$ and $(J, \overline{J})$ be a partition of $\{1, \ldots, n\}$ where $J = \{i_1, \ldots, i_m\} \subseteq \{1, \ldots, n\}$ has m elements.
- We denote by $\mathbf{x}_J = \left(x_{i_1}, \ldots, x_{i_m}\right)$ the restriction of any vector $\mathbf{x} \in \mathbb{F}_q^n$ to the coordinates indexed by J.
- Via the operation of **puncturing** and **shortening** we can obtained codes of shorter lenght from $\mathcal{C}$.

Puncturing a code $(\mathcal{C}_J)$

We can punctured $\mathcal{C}$ by deleting columns from a generator matrix of $\mathcal{C}$ i.e.

$\mathcal{C}_J = \{\mathbf{c}_{\overline{J}} \mid \mathbf{c} \in \mathcal{C}\} \implies \mathcal{C}_J$ is an $[n(\mathcal{C}) - m, k(\mathcal{C}_J), d(\mathcal{C}_J)]$ code with

$$d(\mathcal{C}) - m \leq d(\mathcal{C}_J) \leq d(\mathcal{C}) \quad \text{and} \quad k(\mathcal{C}) - m \leq k(\mathcal{C}_J) \leq k(\mathcal{C})$$

- Moreover if $m < d(\mathcal{C})$ then $k(\mathcal{C}_J) = k(\mathcal{C})$.

Shortening a code $\left(\mathcal{C}^J\right)$

We can shorten $\mathcal{C}$ by deleting columns from a parity check matrix of $\mathcal{C}$. Thus the words of $\mathcal{C}^J$ are codewords of the initial code that have a zero in the J-location, i.e.

$\mathcal{C}^J = \{\mathbf{c}_{\overline{J}} \mid \mathbf{c} \in \mathcal{C} \text{ and } \mathbf{c}_J = 0\} \Rightarrow \mathcal{C}^J$ is an $[n(\mathcal{C}) - m, k(\mathcal{C}^J), d(\mathcal{C}^J)]$ code with

$$d(\mathcal{C}) \leq d(\mathcal{C}^J) \quad \text{and} \quad k(\mathcal{C}) - m \leq k(\mathcal{C}^J) \leq k(\mathcal{C})$$

Modifying Codes

Some properties of these operations

1. $\mathcal{C}^J \subseteq \mathcal{C}_J$.
2. $\dim(\mathcal{C}^J) + \dim\left(\mathcal{C}_{\overline{J}}\right) = \dim(\mathcal{C})$.
3. $(\mathcal{C}_J)^\perp = (\mathcal{C})^J$ and $(\mathcal{C}^J)^\perp = (\mathcal{C}^\perp)_J$.

Lemma 1

Let $\mathcal{C}$ be an MDS code.
If $n(\mathcal{C}) - m \geq k(\mathcal{C})$, then $\mathcal{C}_J$ and $\mathcal{C}^J$ are MDS codes with parameters:

$$[n(\mathcal{C}) - m, k(\mathcal{C})] \quad \text{and} \quad [n(\mathcal{C}) - m, k(\mathcal{C}) - m],$$

respectively.

Generalized Reed-Solomon codes (GRS codes)

Let

- $\mathbf{a} = (a_1, \ldots, a_n)$ be an n-tuple of **mutually distinct** elements of $\mathbb{P}^1(\mathbb{F}_q)$.
- $\mathbf{b} = (b_1, \ldots, b_n)$ be an n-tuple of **nonzero** elements of $\mathbb{F}_q$.

The **GRS** code $\mathrm{GRS}_k(\mathbf{a}, \mathbf{b})$ is defined by:

$$\mathrm{GRS}_k(\mathbf{a}, \mathbf{b}) = \{(f(a_1)b_1, \ldots, f(a_n)b_n) \mid f \in \mathbb{F}_q[X] \text{ and } \deg(f) < k\}$$

Theorem: Parameters of $\mathrm{GRS}_k(\mathbf{a}, \mathbf{b})$

- The $\mathrm{GRS}_k(\mathbf{a}, \mathbf{b})$ is an **MDS** code with parameters $[n, k, n-k+1]$.
- Furthermore a generator matrix of $\mathrm{GRS}_k(\mathbf{a}, \mathbf{b})$ is given by

$$G_{\mathbf{a},\mathbf{b}} = \begin{pmatrix} b_1 & \ldots & b_n \\ b_1 a_1 & \ldots & b_n a_n \\ \vdots & \ddots & \vdots \\ b_1 a_1^{k-1} & \ldots & b_n a_n^{k-1} \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} b_1 & \ldots & b_{n-1} & 0 \\ b_1 a_1 & \ldots & b_{n-1} a_{n-1} & 0 \\ \vdots & \ldots & \vdots & \vdots \\ b_1 a_1^{k-2} & \ldots & b_{n-1} a_{n-1}^{k-2} & 0 \\ b_1 a_1^{k-1} & \ldots & b_{n-1} a_{n-1}^{k-1} & 1 \end{pmatrix}$$

if $a_n = \infty$.

Generalized Reed-Solomon codes (GRS codes)

Proposition GRS

We have

$$\mathrm{GRS}_k(\mathbf{a}, \mathbf{b})^{\perp} = \mathrm{GRS}_{n-k}(\mathbf{a}, \mathbf{s})$$

where $\mathbf{s} = (s_1, \ldots, s_n)$ with $s_i^{-1} = b_i \prod_{j \neq i}(a_i - a_j)$.

Proposition

If $\underline{2 \leq k \leq n-2}$ then a representation of a GRS code is **unique** up to a fractional map of the projective line that induces an automorphism of the code, i.e.

- Different values of $\mathbf{a}$ and $\mathbf{b}$ gives rise to the same GRS code.
- But... the pair $(\mathbf{a}, \mathbf{b})$ is unique up to the action of fractional transformations.

NOTATION

- For all $\mathbf{a}, \mathbf{b} \in \mathbb{F}_q^n$ we define:
  - **Star Multiplication:** $\mathbf{a} * \mathbf{b} = (a_1 b_1, \ldots, a_n b_n) \in \mathbb{F}_q^n$.
  - **Standard Inner Multiplication:** $\mathbf{a} \cdot \mathbf{b} = \sum_{i=1}^{n} a_i b_i$.
- For all subsets $A, \ B \subseteq \mathbb{F}_q^n$ we define:
  - $A * B = \{\mathbf{a} * \mathbf{b} \mid \mathbf{a} \in A \text{ and } \mathbf{b} \in B\}$.
  - $A \perp B \iff \mathbf{a} \cdot \mathbf{b} = 0 \quad \forall\, \mathbf{a} \in A \text{ and } \mathbf{b} \in B$.

Error-correcting pairs (ECP)

Error-correcting pairs (ECP)

Let $\mathcal{C}$ be an $\mathbb{F}_q$ linear code of length n. The pair (A, B) of $\mathbb{F}_{q^N}$-linear codes of length n is a t-ECP for $\mathcal{C}$ over $\mathbb{F}_{q^N}$ if the following properties hold:

- E.1 $(A * B) \perp \mathcal{C}$.
- E.2 $k(A) > t$.
- E.3 $d(B^{\perp}) > t$.
- E.4 $d(A) + d(\mathcal{C}) > n$.

An $[n, k]$ code which has a t-ECP over $\mathbb{F}_{q^N}$ has a decoding algorithm with complexity $\mathcal{O}\left((nN)^3\right)$.

R. Pellikaan

On decoding by error location and dependent sets of error positions.

Discrete Math., 106–107: 369–381 (1992).

R. Kötter.

A unified description of an error locating procedure for linear codes.

In Proceedings of Algebraic and Combinatorial Coding Theory, 113–117. Voneshta Voda (1992).

Examples of the existence of ECP

1. GRS codes

Let

$$A = \mathrm{GRS}_{t+1}(\mathbf{a}, \mathbf{b}_1), \quad B = \mathrm{GRS}_t(\mathbf{a}, \mathbf{b}_2) \quad \text{and} \quad \mathcal{C} = \mathrm{GRS}_{2t}(\mathbf{a}, \mathbf{b}_1 * \mathbf{b}_2)^{\perp}$$

then (A, B) is a t-ECP for $\mathcal{C}$.

Conversely, let $\mathcal{C} = \mathrm{GRS}_k(\mathbf{a}, \mathbf{b})$ then

$$A = \mathrm{GRS}_{t+1}(\mathbf{a}, \mathbf{b}') \quad \text{and} \quad B = \mathrm{GRS}_t(\mathbf{a}, \mathbf{1})$$

is a t-ECP for $\mathcal{C}$ where $t = \left\lfloor \frac{n-k}{2} \right\rfloor$ and $\mathbf{b}' \in (\mathbb{F}_q \setminus \{0\})^n$ verifies that

$$\mathrm{GRS}_k(\mathbf{a}, \mathbf{b})^{\perp} = \mathrm{GRS}_{n-k}(\mathbf{a}, \mathbf{b}').$$

2. Cyclic-codes

I. Duursma

Decoding codes from curves and cyclic codes.
Ph.D thesis, Eindhoven University of Technology (1993)

I. Duursma, R. Kötter.

Error-locating pairs for cyclic codes.
IEEE Trans. Inform. Theory, Vol.40, 1108–1121 (1994)

R. Kötter.

On algebraic decoding of algebraic-geometric and cyclic codes.
Ph.D thesis, Linköping University of Technology (1996).

EXAMPLES OF THE EXISTENCE OF ECP

3. SUBCODES OF A GRS CODE

Let $\mathcal{C}$ be a subcode of a GRS code.

- ➜ This code has an ECP by **Example1** which is also an ECP for $\mathcal{C}$.

4. ALGEBRAIC GEOMETRY CODES

An AG code on a curve of genus g with designed minimum distance d^*:

- ➜ Has a t-ECP over $\mathbb{F}_q$ with $t = \left\lfloor \frac{d^*-1-g}{2} \right\rfloor$.
- ➜ If e is sufficiently large, then there exists a t-ECP over $\mathbb{F}_{q^e}$ with $t = \left\lfloor \frac{d^*-1}{2} \right\rfloor$

R. Pellikaan
On decoding by error location and dependent sets of error positions.
Discrete Math., 106–107: 369–381 (1992).

R. Pellikaan
On the existence of error-correcting pairs.
Statistical Planning and Inference, Vol.51, 229–242. (1996).

5. GOPPA CODES

A Goppa code associated to a Goppa polynomial of degree r can be viewed as an alternant code, i.e. a **subfield subcode** of a GRS code of dimension r.

- ➜ They have an $\left\lfloor \frac{r}{2} \right\rfloor$-ECP.

Properties of ECP

Property 1

If $\mathcal{C}$ is an MDS code and has a t-ECP (A, B) then without loss of generality we may assume that:

- ➜ A is an MDS code with parameters $[n, t+1, n-t]$.
- ➜ B is an MDS code with parameters $[n, t, n-t+1]$.

Property 2

If the property $E.4$ is replaced by the following statements:

E.5 $d(A^{\perp}) > 1$ i.e. A is non-degenerated code.

E.6 $d(A) + 2t > n$.

Then (A, B) is a t-ECP for $\mathcal{C}$ and $d(\mathcal{C}) \geq 2t + 1$.

R. Pellikaan
On decoding by error location and dependent sets of error positions.
Discrete Math., 106–107: 369–381 (1992).

R. Pellikaan
On the existence of error-correcting pairs.
Statistical Planning and Inference, Vol.51, 229–242. (1996).

Motivation: Code-based cryptography is an interesting candidate for Post-Quatum Cryptography

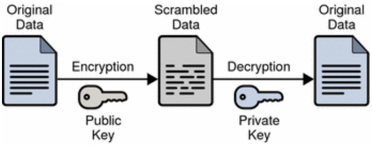
Two keys:

- Private Key: Known only by the recipient.
- Public Key: Available to anyone.

Most PKC are based on number-theoretic problems

→ Quatum computers will break the most popular PKCs: RSA, DSA, ECDSA, ECC, HECC, ...

can be attacked in polynomial time using **Shor's algorithm**



Good news: Post-quatum cryptography

- Hash-based cryptography,
- Code-based cryptography,
- Lattice-based cryptography,
- Multivariate-quadratic-equation cryptography

D. J. Bernstein, J. Buchmann, E. Dahmen.
Post-Quatum Cryptography.
Springer, 2009.

MOTIVATION

"At the heart of any public-key cryptosystem is a one-way function - a function $y = f(x)$ that is easy to evaluate but for which is computationally infeasible (one hopes) to find the inverse $x = f^{-1}(y)$".

N. Koblitz, A. Menezes.
The brave new world of bodacious assumptions in cryptography.
Notices Amer. Math. Soc. 57(3), 357-365 (2010).

Let $\mathcal{C}_t$ the class of linear codes over $\mathbb{F}_q$ that have a t-ECP over an extension of $\mathbb{F}_q$.

- This family have an **efficient decoding algorithm** $\Rightarrow$ they are appropriate for code-based cryptography.
- Most families of codes used in code-based cryptography belongs to $\mathcal{C}_t$.

  (Like GRS codes, Goppa codes, AG codes ...)
- We proposed to use the subclass of $\mathcal{C}_t$ formed by **those linear codes $\mathcal{C}$ whose error correcting pair is not easily reconstructed from** $\mathcal{C}$, i.e. we consider the following one way function:

$$\mathbf{x} = (A, B) \quad \longmapsto \quad \mathbf{y} = A * B,$$

  where (A, B) is a t-ECP.

MOTIVATION: THE CLASS OF GRS CODES WAS PROPOSED FOR CODE-BASED PKC BY NIEDERREITER

- **Sidelnikov-Shestakov** in 1992 introduced an algorithm that breaks the original Niederreiter cryptosystem in polynomial time.
- **Berger and Loidreau** in 2005 propose another version of the Niederreiter scheme designed to resist the Sidelnikov-Shestakov attack.
  - **Main idea:** work with subcodes of the original GRS code.
- **Attacks:**
  1. **Wieschebrink:**
     - Presents the first feasible attack to the Berger-Loidreau cryptosystem but is impractical for small subcodes.
     - Notes that if the square code of a subcode of a GRS code of parameters $[n, k]$ is itself a GRS code of dimension $2k - 1$ then we can apply Sidelnikov-Shestakov attack.
  2. **M-Mártinez-Pellikaan:** Give a characterization of the possible parameters that should be used to avoid attacks on the Berger-Loidreau cryptosystem.

T. Berger and P. Loidreau.
How to mask the structure of codes for a cryptographic use.
Designs, Codes and Cryptography, 35: 63–79, 2005.

I. Márquez-Corbella, E. Martínez-Moro and R. Pellikaan.
The non-gap sequence of a subcode of a generalized Reed-Solomon code.
Proceedings of the Seventh International Workshop on Coding and Cryptography, April 11-15, Paris, France, 183-193, 2011.

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.

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.

V. M. Sidelnikov and S. O. Shestakov.
On insecurity of cryptosystems based on generalized Reed-Solomon codes.
Discrete Mathematics and Applications.

OUR GOAL

THEOREM:

If $\mathcal{C}$ is an MDS code over $\mathbb{F}_q$ of minimum distance $d(\mathcal{C}) = 2t + 1$ and with a t-ECP over a finite extension of $\mathbb{F}_q$ then $\mathcal{C}$ is a GRS code.

WHAT DO WE HAVE?

- In the special cases $k(\mathcal{C}) = \{0, 1, n(\mathcal{C}) - 1, n(\mathcal{C})\}$ the hypothesis of having a t-ECP is not a necessary condition.
  - The $[2t, 0, 2t + 1]$-code is the trivial code $\mathcal{C}_1 = \{\mathbf{0}\}$ which is MDS and $\mathcal{C}_1 = \mathrm{GRS}_0(\mathbf{a}, \mathbf{b})$ for every $\mathbf{a}, \mathbf{b} \in \mathbb{F}_q^{2t}$ that satisfy the right conditions of GRS codes.
  - ➜ The $[2t, 2t, 1]$-code is $\mathcal{C}_1^\perp = \mathbb{F}_q^{2t} = \mathrm{GRS}_{2t}(\mathbf{a}, \mathbf{b}')$ which is MDS, where $\mathbf{b}'$ take the form described in **Proposition GRS**.
  - The $[2t, 1, 2t]$-code is a code $\mathcal{C}_2$ generated by a word $\mathbf{b} \in (\mathbb{F}_q \setminus \{0\})^{2t}$, i.e. $\mathcal{C}_2 = \mathrm{GRS}_1(\mathbf{a}, \mathbf{b})$ for every $\mathbf{a} \in \mathbb{F}_q^{2t}$ that satisfy the right conditions of GRS codes.
  - ➜ If $k(\mathcal{C}) = n - 1$ then its dual $\mathcal{C}^\perp$ belongs to the previous case $\Rightarrow \mathcal{C}$ is a GRS code (using **Proposition GRS**).
- Therefore we need to prove the result for $\underline{2 \leq k(\mathcal{C}) \leq n(\mathcal{C}) - 2}$.
  - When $t = 1$, it is easy to prove that $\mathcal{C}$ is a GRS code.
  - The case $t = 2$ was already proved by Pellikaan.

R. Pellikaan
On the existence of error-correcting pairs.
Statistical Planning and Inference, Vol.51, 229–242. (1996).

  - For $t \geq 2$ ... **Work in progress!!**

➜ If $\mathcal{C}$ has a t-ECP then the code obtained from $\mathcal{C}$ by puncturing twice at any pair of coordinates has a $(t-1)$-ECP.

Thank you for your attention!