Error-correcting pairs and arrays from algebraic geometry codes

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.

Applications of Computer Algebra - ACA 2013

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

$$d(\mathcal{C}) = \min \{ d_H(\mathbf{c}_1, \mathbf{c}_2) \mid \mathbf{c}_1, \mathbf{c}_2 \in \mathcal{C} \text{ and } \mathbf{c}_1 \neq \mathbf{c}_2 \}.$$

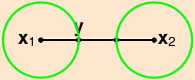


FIGURE: If $d(\mathcal{C}) = 3$

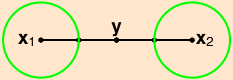


FIGURE: If $d(\mathcal{C}) = 4$

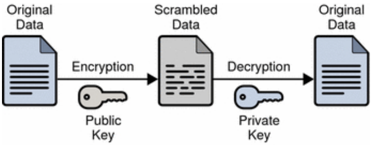
Public-Key Cryptosystems

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.

McEliece cryptosystem

Key Generation

1. Given:
   - $\mathcal{C}$ an $[n, k, d]$ linear code over $\mathbb{F}_q$
   - $G \in \mathbb{F}_q^{k \times n}$ a generator matrix of $\mathcal{C}$.
   - $S \in \mathbb{F}_q^{k \times k}$ a nonsingular matrix.
   - $P \in \mathbb{F}_q^{n \times n}$ a permutation matrix.
2. McEliece Public Key : $(G' = SGP, t)$.
3. McEliece Private Key: (G, S, P)

Encryption

Encrypt a message $\mathbf{m} \in \mathbb{F}_q^k$ as

$$\mathbf{y}' = \mathbf{m}G' + \mathbf{e}'$$

where $\mathbf{e}$ and $e' = eP$ in $\mathbb{F}_q^n$ are random error vectors of weight t.

Decryption

1. Compute $\mathbf{y} = \mathbf{y}'P^{-1} = \mathbf{m}G'P^{-1} + \mathbf{e}'P^{-1} = \mathbf{m}SG + \mathbf{e}$.
2. Apply the decoding algorithm for $\mathcal{C}$ to find $\mathbf{m}S$.
3. $\mathbf{m} = \mathbf{m}SS^{-1}$.

- McEliece introduced the first PKC based on Error-Correcting Codes in 1978.
- Advantages:
  1. Interesting candidate for post-quantum cryptography.
  2. Fast encryption (matrix-vector multiplication) and decryption functions.
- Drawback: Large key size.

R. J. McEliece.
A public-key cryptosystem based on algebraic coding theory.
DSN Progress Report, 42-44:114-116, 1978.

Attacks on the McEliece PKC

➜ Most effective attack against the McEliece cryptosystem is **Information Set Decoding**. Many variants:

1. McEliece (1978)
2. Leon (1988)
3. Lee and Brickell (1988)
4. Stern (1989)
5. van Tilburg (1990)
6. Canteaut and Chabanne (1994)
7. Canteaut and Chabaud (1998)
8. Canteaut and Sendrier (1998)
9. Bernstein, Lange and Peters (2008)
10. Becker, Coron and Joux (2011)

A. Canteaut and H. Chabanne.
A further improvement of the work factor in an attempt at breaking McEliece's cryptosystem.
EUROCODE 94, 1994.

A. Canteaut and F. Chabaud.
A new algorithm for finding minimum-weight words in a linear code: application to McEliece's cryptosystem and to narrow-sense BCH codes of length 511.
IEEE Transaction on Information Theory.

A. Canteaut and N. Sendrier.
Crytanalysis of the original McEliece cryptosystem.
Advances in cryptology - ASIACRYPT'98.

P. J. Lee and E. F. Brickell.
An observation on the security of McEliece's public-key cryptosystem.
Advances in cryptology - EUROCRYPT'98.

A. Becker, J. S. Coron and A. Joux
Improved generic algorithms for hard knapsacks.
Advances in cryptology - EUROCRYPT 2011

D. J. Bernstein, T. Lange, C. Peters.
Attacking and defending the McEliece cryptosystem.
Post-Quantum Cryptography

J. S. Leon.
A probabilistic algorithm for computing minimum weights of large error-correcting codes.
IEEE Transaction on Information Theory.

R. J. McEliece.
A public-key cryptosystem based on algebraic coding theory.
DSN Progress Report

J. Stern.
A method for finding codewords of small weight.
Coding theory and applications, Vol 388 of Lecture Notes in Computer Science, 106-113. Springer, New York, 1989.

J. van Tilburg.
On the McEliece public-key cryptosystem.
Advances in cryptology - CRYPTO'88.

Niederreiter Cryptosystem

- Niederreiter presents a dual version of McEliece cryptosystem in 1986 which is equivalent in terms of security, with the same Goppa codes.

Key Generation

1. Given:
   - $\mathcal{C}$ an $[n, k, d]$ linear code over $\mathbb{F}_q$
   - $H \in \mathbb{F}_q^{(n-k)\times n}$ a parity check matrix of $\mathcal{C}$.
   - $S \in \mathbb{F}_q^{(n-k)\times(n-k)}$ a nonsingular matrix.
   - $P \in \mathbb{F}_q^{n\times n}$ a permutation matrix.
2. Niederreiter Public Key : $(H' = SHP, t)$.
3. Niederreiter Private Key: (H, S, P)

Encryption

Encrypt a message $\mathbf{m} \in \mathbb{F}_q^k$ as

$$\mathbf{y}' = \mathbf{m}H'^T.$$

Decryption

1. Compute $\mathbf{y} = \mathbf{y}' = (S^{-1})^T = \mathbf{m}P^T H^T = \mathbf{m}' H^T$. Syndrome of $\mathbf{m}'$ by H.
2. Apply decoding algorithm for $\mathcal{C}$ to find $\mathbf{m}' = \mathbf{m}P^T$ and thereby $\mathbf{m}$.

- In its original paper Niederreiter proposed the class of GRS codes over $\mathbb{F}_{2^m}$.

H. Niederreiter.
Knapsack-type crypto system and algebraic coding theory.
Problems of Control and Information Theory, 1986.

Y. Xing Li, R. H. Deng and X. Mei Wang.
On the equivalence of McEliece's and Niederreiter public-key cryptosystems.
IEEE Transaction on Information Theory, 1994.

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 $\mathbb{F}_{q^m}$-linear codes of length n is a t-ECP for $\mathcal{C}$ over $\mathbb{F}_{q^m}$ 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 $\mathbb{F}_{q^m}$ has a decoding algorithm with complexity $\mathcal{O}\left((nm)^3\right)$.

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).

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.

Algorithm to find the error positions I

Let

- $\mathcal{C}$ be an $\mathbb{F}_q$-linear code of length n.
- A and B be linear subspaces of $\mathbb{F}_{q^m}^n$.
- $\mathbf{y} \in \mathbb{F}_q^n$ be the received word with error vector $\mathbf{e}$ ➜ $\mathbf{y} = \mathbf{c} + \mathbf{e}$ for some $\mathbf{c} \in \mathcal{C}$

Define:

$$K_{\mathbf{y}} = \{\mathbf{a} \in A \mid \langle \mathbf{y}, \mathbf{a} * \mathbf{b} \rangle = 0, \text{ for all } \mathbf{b} \in B\}$$

Lemma 1:

$$\text{If } A * B \subseteq \mathcal{C}^{\perp} \quad \Longrightarrow \quad K_{\mathbf{y}} = K_{\mathbf{e}}$$

Let J be a subset of $\{1, \ldots, n\}$, define:

$$A(J) = \{\mathbf{a} \in A \mid a_j = 0, \text{ for all } j \in J\}$$

Lemma 2:

$$\text{If } (A * B) \perp \mathcal{C} \text{ and } I = \mathrm{supp}(\mathbf{e}) \quad \Longrightarrow \quad A(I) \subseteq K_{\mathbf{y}} \quad .$$

Moreover, if $d(B^{\perp}) > w_H(\mathbf{e}) \quad \Longrightarrow \quad A(I) = K_{\mathbf{y}}$

Lemma 3:

$$\text{If } k(A) > t \geq w_H(\mathbf{e}) \text{ and } I = \mathrm{supp}(\mathbf{e}) \quad \Longrightarrow \quad \exists \mathbf{a} \in A(I) \setminus \{\mathbf{0}\}$$

Algorithm to find the error positions II

Lemma 4:

If $d(A) + d(\mathcal{C}) > n \quad \Longrightarrow \quad \forall \mathbf{a} \in A \;:\; |J| = n - |\mathrm{supp}(\mathbf{a})| \leq d(\mathcal{C}) - 1$

Moreover, if $d(B^{\perp}) > t \geq \mathrm{w}_H(\mathbf{e}) \quad \Longrightarrow \quad I = \mathrm{supp}(\mathbf{e}) \subseteq J$

1. Compute:
$$K_{\mathbf{y}} = \{\mathbf{a} \in A \mid \langle \mathbf{y}, \mathbf{a} * \mathbf{b} \rangle = 0, \text{ for all } \mathbf{b} \in B\}$$

Find the zero space of a set of linear equations over $\mathbb{F}_{q^m}$

2. If $K_{\mathbf{y}} = \mathbf{0} \Longrightarrow$ **The received word has more than t errors**

➜ Else take a nonzero $\mathbf{a} \in K_{\mathbf{y}} \subseteq A(I)$ and define $J = \{j \mid a_j = 0\}$

3. Find $\mathbf{e} \in \mathbb{F}_q^n$ by solving the following linear equation (which has a **unique** solution):
$$H\mathbf{x}^T = H\mathbf{y}^T \qquad \text{such that} \qquad x_j = 0 \text{ for } j \in J$$

Solve linear equations over $\mathbb{F}_q$

Complexity: $\sim \mathcal{O}\left((nm)^3\right)$

Generalized Reed-Solomon Codes

Let

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

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

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

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

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

➜ A generator matrix of $\mathrm{GRS}_k(\mathbf{a}, \mathbf{b})$ is given by

$$G_{\mathbf{a},\mathbf{b}} = \begin{pmatrix} b_1 & \dots & b_n \\ b_1 a_1 & \dots & b_n a_n \\ \vdots & \ddots & \vdots \\ b_1 a_1^{k-1} & \dots & b_n a_n^{k-1} \end{pmatrix} \in \mathbb{F}_q^{k \times n}$$

t-ECP FOR GRS

Note that: $\mathrm{ev}_{\mathbf{a},\mathbf{b}}(f(X)) * \mathrm{ev}_{\mathbf{a},\mathbf{c}}(g(X)) = \mathrm{ev}_{\mathbf{a},\mathbf{b}}(f(X)g(X)) * \mathbf{c}$. Therefore:
$\mathrm{GRS}_k(\mathbf{a},\mathbf{b}) * \mathrm{GRS}_l(\mathbf{a},\mathbf{c}) = \mathrm{GRS}_{k+l-1}(\mathbf{a},\mathbf{b}*\mathbf{c})$

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 $\mathcal{C} = \mathrm{GRS}_{n-2t}(\mathbf{a},\mathbf{b})$ 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 $\mathbf{b}' \in (\mathbb{F}_q \setminus \{0\})^n$ verifies that

$$\mathcal{C}^\perp = \mathrm{GRS}_{n-2t}(\mathbf{a},\mathbf{b})^\perp = \mathrm{GRS}_{2t}(\mathbf{a},\mathbf{b}').$$

Moreover an $[n, n-2t, 2t+1]$ code that has a t-ECP is a GRS code.

GRS for code-based PKC

- The class of **GRS** codes was proposed by **Niederreiter** for code-based PKC.
- **Sidelnikov and Shestakov** introduced an algorithm that breaks the original Niederreiter cryptosystem in polynomial time.

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

Subcodes

➜ Let:

- $\mathcal{D}$ be a code that has (A, B) as t-ECP.
- $\mathcal{C}$ be a subcode of $\mathcal{D}$

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

➜ The class of **subcodes of GRS** codes was proposed by **Berger-Loidreau** for code-based PKC to resist **Sidelnikov-Shestakov** attack.

➜ For certain parameters, this proposal is not secure.

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.

Algebraic geometry codes

- Let:
  - $\mathcal{X}$ be an algebraic curve of genus g defined over the finite field $\mathbb{F}_q$,
  - $P = (P_1, \dots, P_n)$ be an n-tuple of distinct $\mathbb{F}_q$-rational points on $\mathcal{X}$
  - E be a divisor of $\mathcal{X}$ with $\mathrm{supp}(E) \cap P = \emptyset$ and $\deg(E) = m$.

Space of rational functions associated to E

The space of rational functions associated to E is

$$L(E) = \{f \in \mathbb{F}_q(\mathcal{X}) \mid f = 0 \text{ or } (f) + E \geq 0\}$$

- Since $\mathrm{supp}(E) \cap P = \emptyset$ the following **evaluation map** is well defined:

$$\begin{array}{rccc} \mathrm{ev}_P : & L(E) & \longrightarrow & \mathbb{F}_q^n \\ & f & \longmapsto & \mathrm{ev}_P(f) = (f(P_1), \dots, f(P_n)) \end{array}$$

Riemman-Roch Theorem

$$\dim L(E) \geq m + 1 - g.$$

Furthermore if $m > 2g - 2$ then $\dim L(E) = m + 1 - g$.

ALGEBRAIC GEOMETRY CODES

ALGEBRAIC GEOMETRY CODES (AG CODES)

The AG code associated to $\mathcal{X}$, $P = (P_1, \ldots, P_n)$ and E is

$$\mathcal{C}_L(\mathcal{X}, P, E) = \{\mathrm{ev}_P(f) \mid f \in L(E)\}$$

THEOREM: PARAMETERS OF AN AG CODE

If $n > m$ then $\mathcal{C}_L(\mathcal{X}, P, E)$ is an $[n, k, d]$ code over $\mathbb{F}_q$ where

$$k \geq m + 1 - g \quad \text{and} \quad d \geq n - m$$

Moreover, if $m > 2g - 2$ then $k = m + 1 - g$.

- If $\{f_1, \ldots, f_k\}$ is a basis of $L(E)$ then

$$G = \begin{pmatrix} f_1(P_1) & \ldots & f_1(P_n) \\ \vdots & & \vdots \\ f_k(P_1) & \ldots & f_k(P_n) \end{pmatrix} \in \mathbb{F}_q^{k \times n}$$

is a generator matrix of the code $\mathcal{C}_L(\mathcal{X}, P, E)$

t-ECP FOR AG CODES I

Let F and G be divisors of $\mathcal{X}$. Then there is a well defined linear map:

$$\begin{array}{ccc} L(E) \otimes L(G) & \longrightarrow & L(F+G) \\ f \otimes g & \longmapsto & fg \end{array}$$

Hence:

$$\mathcal{C}_L(\mathcal{X},\mathcal{P},F) * \mathcal{C}_L(\mathcal{X},\mathcal{P},G) \subseteq \mathcal{C}_L(\mathcal{X},\mathcal{P},F+G)$$

Let $\mathcal{C} = \mathcal{C}_L(\mathcal{X},\mathcal{P},E)^{\perp}$ and choose a divisor F with disjoint support from $\mathcal{P}$, then:

$$A = \mathcal{C}_L(\mathcal{X},\mathcal{P},E) \quad \text{and} \quad B = \mathcal{C}_L(\mathcal{X},\mathcal{P},E-F)$$

is a t-ECP for $\mathcal{C}$.

t-ECP for AG codes II

Proposition: [Pellikaan (1989)]

An AG code on a curve of genus g with designed minimum distance d has a t-ECP over $\mathbb{F}_q$ with

$$t = \left\lfloor \frac{d-1-g}{2} \right\rfloor$$

Proposition: [Pellikaan (1996)]

If m is sufficiently large then an AG code with designed minimum distance d has a t-ECP over $\mathbb{F}_{q^m}$ where

$$t = \left\lfloor \frac{d-1}{2} \right\rfloor$$

- ➜ With **ECPs** we do not have a constructive decoding scheme
- ➜ **BUT** an **Error-correcting array** gives a decoding algorithm that decodes up to

$$\left\lfloor \frac{d^*-1}{2} \right\rfloor \text{ where } d^* = \text{Feng-Rao designed minimum distance}$$

with complexity $\mathcal{O}\left(n^3\right) \Longrightarrow$ **Majority coset decoding**

AG codes for code-based PKC

- In 1996 **Janwa and Moreno** propose to use AG codes for the McEliece cryptosystem.
- This system was broken for:
  1. Codes on curves of genus $g = 0$ by the **Sidelnikov-Shestakov** attack.

     GRS codes are Algebraic Geometry codes on the projective line.
  2. Codes on curves of genus $g \leq 2$ by **Faure and Minder**.
  3. **VSAG** are not secure for McEliece cryptosystem by **M-Martínez-Pellikaan-Ruano**

Very Strong Algebraic-Geometric (VSAG) codes

A code $\mathcal{C}$ has a VSAG representation if $\mathcal{C} = \mathcal{C}_L(\mathcal{X}, P, E)$ where the curve $\mathcal{X}$ has genus g, P consists of n points and E has degree m such that

$$2g + 2 < m < \tfrac{1}{2}n \quad \text{or} \quad \tfrac{1}{2}n + 2g - 2 < m < n - 4$$

C. Faure and L. Minder.
Cryptanalysis of the McEliece cryptosystem over hyperelliptic codes.
Proceedings 11th Int. Workshop on Algebraic and Combinatorial Coding Theory, 2008.

H. Janwa and O. Moreno.
McEliece public crypto system using algebraic-geometric codes.
Designs, Codes and Cryptography, 1996.

I. Márquez-Corbella, E. Martínez-Moro and R. Pellikaan.
On the unique representation of very strong algebraic geometry codes.
Designs, Codes and Cryptography, 2013.

I. Márquez-Corbella, E. Martínez-Moro, R. Pellikaan and D. Ruano.
Computational aspects of retrieving a representation of an algebraic geometry code.
Submitted to Designs, Codes and Cryptography.

Alternant codes

Let

- $\mathbf{a} = (a_1, \ldots, a_n)$ be an n-tuple of **mutually distinct** elements of $\mathbb{F}_{q^m}$.
- $\mathbf{b} = (b_1, \ldots, b_n)$ be an n-tuple of **nonzero** elements of $\mathbb{F}_{q^m}$.
- → $\mathrm{GRS}_k(\mathbf{a}, \mathbf{b})$ be the GRS code over $\mathbb{F}_{q^m}$ of dimension k.

The **alternant code** $\mathrm{Alt}_r(\mathbf{a}, \mathbf{b})$ is the $\mathbb{F}_q$-linear restriction:

$$\mathrm{Alt}_r(\mathbf{a}, \mathbf{b}) = \mathbb{F}_q^n \cap (\mathrm{GRS}_r(\mathbf{a}, \mathbf{b}))^\perp$$

Parameters of $\mathrm{Alt}_r(\mathbf{a}, \mathbf{b})$

The $\mathrm{Alt}_r(\mathbf{a}, \mathbf{b})$ has parameters $[n, k, d]_q$ with:

$$k \geq n - mr \qquad \text{and } d \geq r + 1$$

Every $[n, k, d]$ linear code with $d \geq 2$ is an **alternant code**!

t-ECP for Alternant codes

Let $\mathcal{C} = \mathrm{Alt}_{2t}(\mathbf{a}, \mathbf{b})$. Then:

$$d(\mathcal{C}) \geq 2t + 1 \quad \text{and} \quad \mathcal{C} \subseteq (\mathrm{GRS}_{2t+1}(\mathbf{a}, \mathbf{b}))^{\perp}$$

Let

$$A = \mathrm{GRS}_{t+1}(\mathbf{a}, \mathbf{1}), \quad \text{and} \quad B = \mathrm{GRS}_t(\mathbf{a}, \mathbf{b})$$

then (A, B) is a t-ECP over $\mathbb{F}_{q^m}$ for $\mathcal{C}$.

No known structural attacks against code-base PKC using Alternant codes

Goppa codes

Let

- $\mathbf{a} = (a_1, \ldots, a_n)$ be an n-tuple of **mutually distinct** elements of $\mathbb{F}_{q^m}$.
- g be a polynomial **with coefficients** in $\mathbb{F}_{q^m}$ such that

$$g(a_j) \neq 0 \text{ for all } j = 1, \ldots, n$$

The **Goppa code** $\Gamma(\mathbf{a}, g)$ is the $\mathbb{F}_q$-linear code defined by:

$$\Gamma(\mathbf{a}, g) = \left\{ \mathbf{c} \in \mathbb{F}_q^n \mid \sum_{j=1}^{n} \frac{c_j}{X - a_j} \equiv 0 \mod g(X) \right\}$$

Goppa codes are alternant codes

Let

- $\mathbf{a} = (a_1, \ldots, a_n)$ be an n-tuple of **mutually distinct** elements of $\mathbb{F}_{q^m}$.
- g be a **Goppa polynomial** of degree r.
- $\mathbf{b} = (b_1, \ldots, b_n)$ be an n-tuple of **nonzero** elements of $\mathbb{F}_{q^m}$ such that $b_j = \frac{1}{g(a_j)}$

Then: $\Gamma(\mathbf{a}, g) = \mathrm{Alt}_r(\mathbf{a}, \mathbf{b}) \Longrightarrow$ it has an $\lfloor \frac{r}{2} \rfloor$-ECP

Binary Goppa Codes

Let:

- $\mathbf{a} = (a_1, \ldots, a_n)$ be an n-tuple of **mutually distinct** elements of $\mathbb{F}_{2^m}$.
- g be a Goppa polynomial with coefficients in $\mathbb{F}_{2^m}$ of **degree r**.
- ➜ And suppose moreover that g has **no square factor**.

Then:

1. $\Gamma(\mathbf{a}, g) = \Gamma(\mathbf{a}, g^2)$.
2. $\Gamma(\mathbf{a}, g)$ has parameters $[n, k, d]$ with

$$k \geq n - mr \qquad \text{and} \qquad d \geq 2r + 1$$

$\Gamma(\mathbf{a}, g)$ has an r-ECP

BCH codes

Let

- $n = q^m - 1$
- α be a primitive element of $\mathbb{F}^*_{q^m}$
- $m_i(X)$ be the minimal polynomial of α^i

The **primitive narrow-sense BCH code** over $\mathbb{F}_q$ of length n and distance at least d is the **cyclic code** with generator polynomial: $g(X) = \mathrm{lcm}\,(m_i(X), \ldots, m_{d-1}(X))$

BCH Codes are alternant codes

Let

- $\mathbf{a} = (a_1, \ldots, a_n)$ with $a_i = \alpha^{i-1}$ for $i = 1, \ldots, n$.
- $\mathbf{b} = \mathbf{1} \in \mathbb{F}^n_{q^m}$.

Then the BCH code with defining zeros $\mathcal{Z} = \{0, 1, \ldots, \delta - 2\}$ is: $\mathrm{Alt}_{\delta-1}(\mathbf{a}, \mathbf{b})$.

➜ ECP for cyclic codes were found **beyond half the BCH bound** by Duursma and Kötter.

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).

Structural attacks against Code-based PKC I

GRS codes

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

Subcodes of GRS codes

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.

Structural attacks against Code-based PKC II

AG codes

C. Faure and L. Minder.
Cryptanalysis of the McEliece cryptosystem over hyperelliptic codes.
Proceedings 11th Int. Workshop on Algebraic and Combinatorial Coding Theory, 2008.

I. Márquez-Corbella, E. Martínez-Moro and R. Pellikaan.
On the unique representation of very strong algebraic geometry codes.
Designs, Codes and Cryptography, 2013.

I. Márquez-Corbella, E. Martínez-Moro, R. Pellikaan and D. Ruano.
Computational aspects of retrieving a representation of an algebraic geometry code.
Submitted to Designs, Codes and Cryptography.

Subfield subcodes of GRS codes = Alternant codes:
OPEN

QUESTION:

➜ If a code has a ***t*-ECP** how difficult / easy is to **retrieve** such a pair?

Reverse engineering AG codes

Let

- $\mathcal{X}$ be an algebraic curve over $\mathbb{F}_q$ of genus g.
- E be a divisor of $\mathcal{X}$ with $\deg(E) = m$ and $\{f_0, \ldots, f_r\}$ be a basis of $L(E)$.

We consider the following map:

$$\begin{array}{cccc} \varphi_E : & \mathcal{X} & \longrightarrow & \mathbb{P}^r(\mathbb{F}_q) \\ & P & \longmapsto & \varphi_E(P) = (f_0(P), \ldots, f_r(P)) \end{array}$$

Curves defined by quadratic equations

1. If $m \geq 2g + 2$ then $\varphi_E(\mathcal{X}) = \mathcal{Y}$ is a normal curve in $\mathbb{P}^{m-g}$ which is the intersection of quadrics.

➜ In particular $I(\mathcal{Y})$ is generated by $I_2(\mathcal{Y})$.

2. If $m \geq 2g + 1$ then $\varphi_E(\mathcal{X}) = \mathcal{Y}$ is a normal curve in $\mathbb{P}^{m-g}$ which is the intersection of quadrics and cubics.

➜ In particular $I(\mathcal{Y})$ is generated by $I_2(\mathcal{Y})$ and $I_3(\mathcal{Y})$.

D. Mumford.
Varieties defined by quadratic equations.
Questions on algebraic varieties, C.I.M.E, III Ciclo, Varenna, 1969, pp. 29–100, Edizioni Cremonese, Rome, 1970.

B. Saint-Donat.
Sur les équatons définissant une courbe algébrique.
C. R. Acad. Sci. Paris Sr. A, volume 274, pp. 487–489, 1972.

$I_d(\mathcal{Y})$

$I_d(\mathcal{Y})$ is the ideal generated by the homogeneous elements of degree d in $I(\mathcal{Y})$.

REVERSE ENGINEERING AG CODES II

Let

- $\mathcal{Y}$ be an algebraic curve in $\mathbb{P}^r$ of degree m such that $I(\mathcal{Y}) = I_2(\mathcal{Y})$.
- Q be an n-tuples of points that lies on the curve $\mathcal{Y}$ (i.e. $I(\mathcal{Y}) \subseteq I(Q)$)

PROPOSITION: DUE TO BEZOUT-THEOREM

If $n > 2m$ then,

$$I_2(Q) = I_2(\mathcal{Y})$$

Reverse engineering AG codes III

➜ Let $\mathcal{C}$ be a k-dimensional subspace of $\mathbb{F}^n$ with basis $\{g_1, \ldots, g_k\}$.

We denote:

1. **The second symmetric power of $\mathcal{C}$** by $S^2(\mathcal{C})$
   - $S^2(\mathcal{C})$ has basis $\{X_iX_j \mid 1 \leq i \leq j \leq n\}$ and dimension $\binom{k+1}{2}$
2. **The square code of $\mathcal{C}$** by $\langle \mathcal{C} * \mathcal{C} \rangle$ or by $\mathcal{C}^{(2)}$.
   - $\mathcal{C}^{(2)}$ is the linear subspace in $\mathbb{F}^n$ generated by $\{\mathbf{a} * \mathbf{b} \mid \mathbf{a}, \mathbf{b} \in \mathcal{C}\}$.

➜ We consider the linear map:

$$\begin{array}{cccc} \sigma : & S^2(\mathcal{C}) & \longrightarrow & \mathcal{C}^{(2)} \\ & X_iX_j & \longmapsto & g_i * g_j \end{array}$$

We denote by $K_2(\mathcal{C})$ the kernel of this map, then

$$0 \longrightarrow K_2(\mathcal{C}) \longrightarrow S^2(\mathcal{C}) \longrightarrow \mathcal{C}^{(2)} \longrightarrow 0$$

is an **exact sequence**.

And

$$I_2(Q) = \left\{ \sum_{1 \leq i \leq j \leq k} a_{ij} X_i X_j \mid \sum_{1 \leq i \leq j \leq k} a_{ij} \left(g_i * g_j\right) = 0 \right\} = K_2(\mathcal{C})$$

Proposition

Let Q be an n-tuple of points in $\mathbb{P}^r$ over $\mathbb{F}$ not in a hyperplane. Then the complexity of the computation of $I_2(Q)$ is at most $\mathcal{O}\left(n^2 \binom{r}{2}\right)$.

Reverse engineering AG codes IV

Very Strong Algebraic-Geometric (VSAG) codes

A code $\mathcal{C}$ has a VSAG representation if $\mathcal{C} = \mathcal{C}_L(\mathcal{X}, P, E)$ where the curve $\mathcal{X}$ has genus g, P consists of n points and E has degree m such that

$$2g + 2 < m < \tfrac{1}{2}n \quad \text{or} \quad \tfrac{1}{2}n + 2g - 2 < m < n - 4$$

The **dual** of a VSAG code is again VSAG

- The dimension of such a code is $k = m + 1 - g$. Thus the dimension satisfies the following bound:

$$g + 3 < k < \tfrac{1}{2}n - g + 1 \quad \text{or} \quad \tfrac{1}{2}n + g - 1 < k < n - g - 2$$

Theorem

Let $\mathcal{C}$ be a VSAG code then a VSAG representation can be obtained from its generator matrix.

- Moreover all VSAG representations of $\mathcal{C}$ are strict isomorphic.

Reverse engineering AG codes V

Let $\mathcal{C} = \mathcal{C}_L(\mathcal{X}, P, E)^{\perp}$ be an **AG code** where

- the curve $\mathcal{X}$ has **genus** g.
- the divisor E has degree m such that $2g + 2 \leq m < \frac{1}{2}n$

➜ $\mathcal{C}$ is a **VSAG** code with $d^* = m - 2g + 1$

Our goal: Construct ECP'S eluding the use of Riemann-Roch spaces!!

The **attackers** will use an equivalent representation $(\mathcal{Y}, Q, F)$ of the same code $\mathcal{C}^{\perp}$, which is also **VSAG**.

Reverse engineering AG codes V

Recall that: Let $\mathcal{C} = \mathcal{C}_L(\mathcal{Y}, Q, F)^{\perp}$ be an AG code of **genus** g and **designed minimum distance** d^* such that

$$m = \deg(F) > 2g - 2.$$

Then $\mathcal{C}$ has parameters $[n, \geq n + g - m - 1, \geq m - 2g + 2]$.

Define $A = \mathcal{C}_L(\mathcal{Y}, Q, F - D)$ and $B = \mathcal{C}_L(\mathcal{Y}, Q, D)$

Then $< A * B > \subseteq \mathcal{C}^{\perp}$. Moreover if

$$t = \left\lfloor \frac{d^* - 1 - g}{2} \right\rfloor \quad \text{and} \quad \deg(D) = t + g$$

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

In particular we take $D = (t + g)Q_1$ where Q_1 is the first rational point of Q.

Reverse engineering AG codes VI

Define:

- $A_0 = \mathcal{C}_L(\mathcal{Y}, Q, F - (m - t - g)Q_1)$.
  - Note that $L(F - (m - t - g)Q_1) \subseteq L(F)$, thus $A_0 \subseteq \mathcal{C}^\perp = \mathcal{C}_L(\mathcal{Y}, Q, F)$
  - A_0 is the space of those codewords in $\mathcal{C}^\perp$ that are zero at the first position of multiplicity $m - t - g$. **This multiplicity can be controlled!!!**

- $B_0 = \langle A * \mathcal{C} \rangle^\perp$
  - Note that $B_0^\perp \subseteq B^\perp$ so $d(B_0^\perp) \geq d(B^\perp) > t$

Thus (A_0, B_0) is a t-ECP for $\mathcal{C}$.

Thank you for your attention!