Error-correcting pairs and majority coset decoding for and from algebraic geometry codes

Ruud Pellikaan
g.r.pellikaan@tue.nl
joint work with
Irene Márquez-Corbella

Valladolid, 10 July 2013

Introduction and content

- Error-correcting codes
- Codes on curves
- Error-correcting pairs for codes on curves
- Majority coset decoding
- Public-key cryptosystems
- Code-based cryptography
- Reverse engineering AG codes
- Error-correcting pairs from codes on curves
- Questions

Error-correcting codes

C linear block code: $\mathbb{F}_q$-linear subspace of $\mathbb{F}_q^n$

parameters $[n, k, d]$:
n = length
k = dimension of C
d = minimum distance of C

$$d = \min |\{d(\mathbf{x}, \mathbf{y}) \mid \mathbf{x}, \mathbf{y} \in C, \mathbf{x} \neq \mathbf{y}\}|$$

t = error-correcting capacity of C

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

Inner and star product

The standard inner product is defined by

$$\mathbf{a} \cdot \mathbf{b} = a_1 b_1 + \cdots + a_n b_n$$

For two subsets A and B of $\mathbb{F}_q^n$
$A \perp B$ if and only if $\mathbf{a} \cdot \mathbf{b} = 0$ for all $\mathbf{a} \in A$ and $\mathbf{b} \in B$

Let $\mathbf{a}$ and $\mathbf{b}$ in $\mathbb{F}_q^n$
The star product is defined by coordinatewise multiplication:

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

For two subsets A and B of $\mathbb{F}_q^n$

$$A * B = \{\mathbf{a} * \mathbf{b} \mid \mathbf{a} \in A \text{ and } \mathbf{b} \in B\}$$

Error-correcting pairs - 1

Let C be a linear code in $\mathbb{F}_q^n$

The pair (A, B) of linear subcodes of $\mathbb{F}_{q^m}^n$ is a called a t-error correcting pair (ECP) over $\mathbb{F}_{q^m}$ for C if

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

Error-correcting pairs - 2

Let A and B be linear subspaces of $\mathbb{F}_{q^m}^n$
Let $\mathbf{r} \in \mathbb{F}_q^n$ be a received word
Define the kernel

$$K(\mathbf{r}) = \{\, \mathbf{a} \in A \mid (\mathbf{a} * \mathbf{b}) \cdot \mathbf{r} = 0 \text{ for all } \mathbf{b} \in B \}$$

Lemma

Let C be an $\mathbb{F}_q$-linear code of length n
Let $\mathbf{r}$ be a received word with error vector $\mathbf{e}$
So $\mathbf{r} = \mathbf{c} + \mathbf{e}$ for some $\mathbf{c} \in C$
If $A * B \subseteq C^{\perp}$, then

$$K(\mathbf{r}) = K(\mathbf{e})$$

Error-correcting pairs - 3

Let (A, B) be a t-ECP for C
Let J be a subset of $\{1, \ldots, n\}$
Define the subspace of A

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

Lemma
Let $(A * B) \perp C$
Let $\mathbf{e}$ be an error vector of the received word $\mathbf{r}$
If $I = \text{supp}(\mathbf{e}) = \{ i \mid e_i \neq 0 \}$, then

$$A(I) \subseteq K(\mathbf{r})$$

If moreover $d(B^{\perp}) > \text{wt}(\mathbf{e})$, then $A(I) = K(\mathbf{r})$

Error-correcting pairs - 4

Theorem

Let C be an $\mathbb{F}_q$-linear code of length n
Let (A, B) be a t-error-correcting pair over $\mathbb{F}_{q^m}$ for C

Then the basic algorithm corrects t errors
for the code C with complexity $\mathcal{O}((mn)^3)$

Let $\mathcal{X}$ be an algebraic curve defined over $\mathbb{F}_q$ of genus g
Let $\mathcal{P} = (P_1, \ldots, P_n)$ an n-tuple of mutual distinct points of $\mathcal{X}(\mathbb{F}_q)$

(If the support of E is disjoint from $\mathcal{P}$), then the evaluation map

$$\mathrm{ev}_{\mathcal{P}} : L(E) \to \mathbb{F}_q^n$$

where $\mathrm{ev}_{\mathcal{P}}(f) = (f(P_1), \ldots, f(P_n))$, is well defined.

The algebraic geometry code $C_L(\mathcal{X}, \mathcal{P}, E)$
is the image of $L(E)$ under the evaluation map $\mathrm{ev}_{\mathcal{P}}$
If $m < n$, then $C_L(\mathcal{X}, \mathcal{P}, E)$ is an $[n, k, d]$ code with

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

$n - m$ is called the designed minimum distance of $C_L(\mathcal{X}, \mathcal{P}, E)$

Embedding of $\mathcal{X}$ in linear system of E of degree m

Let $f_1, f_2, \ldots, f_k$ be a basis of $L(E)$

$$\varphi_E : \mathcal{X} \longrightarrow \mathbb{P}^{k-1}$$

$$P \mapsto (f_1(P) : f_2(P) : \ldots : f_k(P))$$

$\mathcal{Y} = \varphi_E(\mathcal{X})$ is a curve of degree m in $\mathbb{P}^{k-1}$

$\mathcal{Q} = (\varphi_E(P_1), \ldots, \varphi_E(P_n))$ projective system

$$G_{\mathcal{Q}} = \begin{pmatrix} f_1(P_1) & \cdots & f_1(P_j) & \cdots & f_1(P_n) \\ f_2(P_1) & \cdots & f_2(P_j) & \cdots & f_2(P_n) \\ \vdots & \ldots & \vdots & \ldots & \vdots \\ f_k(P_1) & \cdots & f_k(P_j) & \cdots & f_k(P_n) \end{pmatrix}$$ generator matrix

minimum distance $\geq n - m$

Let ω be a differential form with
a simple pole at P_j with residue 1 for all $j = 1, \ldots, n$

Let K be the canonical divisor of ω
Let m be the degree of the divisor E on $\mathcal{X}$
with disjoint support from $\mathcal{P}$

Let $E^\perp = D - E + K$ and $m^\perp = \deg(E^\perp)$
Then $m^\perp = 2g - 2 - m + n$ and

$$C_L(\mathcal{X}, \mathcal{P}, E)^\perp = C_L(\mathcal{X}, \mathcal{P}, E^\perp)$$

ECP for AG codes - 1

Let F and G be divisors
Then there is a well defined linear map

$$L(F) \otimes L(G) \longrightarrow L(F+G)$$

given on generators by

$$f \otimes g \mapsto fg$$

Hence

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

ECP for AG codes - 2

Let $C = C_L(\mathcal{X}, \mathcal{P}, E)^{\perp}$

Choose a divisor F with support disjoint from $\mathcal{P}$
Let $A = C_L(\mathcal{X}, \mathcal{P}, F)$
Let $B = C_L(\mathcal{X}, \mathcal{P}, E - F)$

Then
- $A * B \subseteq C^{\perp}$
- If $t + g \leq \deg(F) < n$, then $k(A) > t$
- If $\deg(G - F) > t + 2g - 2$, then $d(B^{\perp}) > t$
- If $\deg(G - F) > 2g - 2$, then $d(A) + d(C) > n$

Proposition

An algebraic geometry code of
designed minimum distance d
from a curve over $\mathbb{F}_q$ of genus g
has a t-error-correcting pair over $\mathbb{F}_q$ where

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

Proposition

An algebraic geometry code of
designed minimum distance d
from a curve over $\mathbb{F}_q$ of genus g
has a t-error-correcting pair over $\mathbb{F}_{q^m}$ where

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

if

$$m > \log_q \left(2\binom{n}{t} + 2\binom{n}{t+1} + 1 \right)$$

Not constructive!
Majority coset decoding gives a constructive and efficient approach

Majority coset decoding - 1

Proposition (Feng-Rao, Duursma)

Let C be code with a subcode D of codimension one
Let $\mathbf{a}_1, \ldots, \mathbf{a}_w$ and $\mathbf{b}_1, \ldots, \mathbf{b}_w$ such that

$$\begin{cases} \mathbf{a}_i * \mathbf{b}_j \in C^\perp & \text{if } i+j \leq w, \\ \mathbf{a}_i * \mathbf{b}_j \in D^\perp \setminus C^\perp & \text{if } i+j = w+1. \end{cases}$$

Then all words of $C \setminus D$ have weight at least w

Majority coset decoding - 2

Proof: Let $\mathbf{c} \in C \setminus D$
Let A be the $w \times n$ matrix with the $\mathbf{a}_i$'s as rows
Let B be the $w \times n$ matrix with the $\mathbf{b}_j$'s as rows
Let $D(\mathbf{c})$ be the diagonal matrix with $\mathbf{c}$ on the diagonal
Let $S(\mathbf{c})$ be the $w \times w$ matrix with entries $s_{i,j} = \mathbf{a}_i * \mathbf{b}_j \cdot \mathbf{c}$
Then

$$AD(\mathbf{c})B^T = S(\mathbf{c})$$

and

$$\begin{cases} s_{i,j} = 0 & \text{if } i + j \leq w, \\ s_{i,j} \neq 0 & \text{if } i + j = w + 1. \end{cases}$$

Hence $\text{wt}(\mathbf{c}) = \text{rk}(D(\mathbf{c})) \geq \text{rk}(S(\mathbf{c})) = w$

Majority coset decoding - 3

Let $w = 2t + 1$

$$\begin{pmatrix}
0 & 0 & \cdots & 0 & 0 & \cdots & 0 & s_{1,w} \\
0 & 0 & \cdots & 0 & 0 & \cdots & s_{2,w-1} & s_{2,w-1} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & 0 & s_{t,t+1} & \cdots & s_{t,w-1} & s_{t,w} \\
0 & 0 & \cdots & s_{t+1,t} & s_{t+1,t+1} & \cdots & s_{t+1,w-1} & s_{t+1,w} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & s_{w-1,2} & \cdots & s_{w-1,t} & s_{w-1,t+1} & \cdots & s_{w-1,w-1} & s_{w-1,w} \\
s_{w,1} & s_{w,2} & \cdots & s_{w,t} & s_{w,t+1} & \cdots & s_{w,w-1} & s_{w,w}
\end{pmatrix}$$

Majority coset decoding - 4

Decoding: Let r be a received word
with $\mathbf{r} = \mathbf{c} + \mathbf{e}$ and $\mathbf{c} \in C \setminus D$ and error vector $\mathbf{e}$
Let $S(\mathbf{r})$ be the $t \times t$ syndrome matrix with entries $s_{i,j}(\mathbf{r}) = \mathbf{a}_i * \mathbf{b}_j \cdot \mathbf{r}$
Then

$$s_{i,j}(\mathbf{r}) = s_{i,j}(\mathbf{e}) \quad \text{if} \quad i + j \leq w$$

are called the known syndromes
Now D has codimension one in C, so there exists a $\mathbf{d} \in D^\perp \setminus C^\perp$
and $\lambda_{ij} \in \mathbb{F}_q^*$ for $i + j = w + 1$ such that

$$\mathbf{a}_i * \mathbf{b}_j \equiv \lambda_{ij}\mathbf{d} \mod C^\perp$$

Hence the unknown syndromes are related to $\mathbf{d} \cdot \mathbf{r}$ by:

$$s_{i,j}(\mathbf{r}) = \lambda_{ij}\mathbf{d} \cdot \mathbf{r} \quad \text{if} \quad i + j = w + 1$$

Majority coset decoding - 5

Let $w = 2t + 1$

$$\begin{pmatrix} s_{1,1} & s_{1,2} & \cdots & s_{1,t} & s_{1,t+1} & \cdots & s_{1,w-1} & s_{1,w} \\ s_{2,1} & s_{2,2} & \cdots & s_{2,t} & s_{2,t+1} & \cdots & s_{2,w-1} & \\ \vdots & \vdots & \ddots & \vdots & \vdots & & & \\ s_{t,1} & s_{t,2} & \cdots & s_{t,t} & s_{t,t+1} & & & \\ s_{t+1,1} & s_{t+1,2} & \cdots & s_{t+1,t} & & & & \\ \vdots & \vdots & & & & & & \\ s_{w-1,1} & s_{w-1,2} & & & & & & \\ s_{w,1} & & & & & & & \end{pmatrix}$$

Public-key cryptosystems - 1

At the heart of any public-key cryptosystem is a one-way function

$$y = f(x)$$

that is easy to evaluate but for which it is computationally infeasible (one hopes) to find an inverse

$$x \in f^{-1}(y)$$

Public-key cryptosystems - 2

PKC systems use trapdoor one-way functions

by mathematical problems that are (supposedly) hard

– RSA, factoring integers
– Diffie-Hellman, discrete-log problem in finite field
– Elliptic curve PKC, addition on elliptic curve
– Lattice-based PKC systems closest vector problem
– Code-based PKC systems, decoding of codes

Code based PKC systems - 1

Code based PKC systems

Take a class of codes that have an efficient decoding algorithm:
Scramble a generator matrix such that it looks like a random code

– Goppa codes (McEliece)
– with parity check matrix instead of generator matrix (Niederreiter)
– Algebraic geometry codes (Janwa-Moreno)
– subcodes of GRS codes (Berger-Loidreau)
– subfield subcodes of algebraic geometry codes (Janwa-Moreno)

Code based PKC systems - 2

Generic attack – decoding algorithms:

– McEliece 1978

.....

– Finiasz-Sendrier 2009

– Bernstein-Lange-Peters 2008-2011

– Becker-Joux-May-Meurer Eurocrypt 2012

Structural attack:

– GRS codes (Sidelnikov-Shestakov)

– subcodes of GRS codes (Wieschebrink, Márquez-Martínez-P)

– Alternant codes: open

– Goppa codes: open

– AG codeds: (Faure-Minder, $g \leq 2$)

– VSAG codes: (Márquez-Martínez-P-Ruano, arbitrary g)

Reverse engineering AG codes - 1

Let $\mathcal{X}$ be an absolutely irreducible and nonsingular curve of genus g over the perfect field $\mathbb{F}$

Let E be a divisor on $\mathcal{X}$ of degree m

If $m \geq 2g + 1$
then φ_E gives an embedding of $\mathcal{X}$ onto $\mathcal{Y} = \varphi_E(\mathcal{X})$
which is a normal curve in the linear system $|E| = \mathbb{P}^{m-g}$

If $m \geq 2g + 2$, then $\mathcal{Y}$ is an intersection of quadrics
More precisely:
$I(\mathcal{Y})$ is generated by $I_2(\mathcal{Y})$
the set of homogeneous elements of degree two in $I(\mathcal{Y})$

Technische Universiteit Eindhoven University of Technology

Reverse engineering AG codes - 2

Let $\mathcal{Y}$ be a curve embedded in projective r-space of degree m
Let $I(\mathcal{Y})$ be the vanishing ideal of $\mathcal{Y}$
Let $\mathcal{Q}$ be a subset of $\mathcal{Y}$ of n points
Then

$$I(\mathcal{Y}) \subseteq I(\mathcal{Q})$$

Hence

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

Suppose $I(\mathcal{Y})$ is generated by $I_2(\mathcal{Y})$

$$\text{If } n > 2m, \text{ then } I_2(\mathcal{Y}) = I_2(\mathcal{Q})$$

By Bézout's Theorem

$\mathbf{g}_1, \ldots, \mathbf{g}_k$ a basis of C

$S^2(C)$ is the second symmetric power of C
$S^2(C)$ has basis $\left\{X_i X_j \mid 1 \le i \le j \le n\right\}$ and dimension $\binom{k+1}{2}$
with $X_i = \mathbf{g}_i$

$C^{(2)} = \langle C * C \rangle$ the square of C

Consider the linear map

$$\begin{array}{rccc} \sigma : & S^2(C) & \longrightarrow & C^{(2)} \\ & x_i X_j & \longmapsto & \mathbf{g}_i * \mathbf{g}_j \end{array}$$

$K_2(C)$ is the kernel of this map

Then

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

is an exact sequence and

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

Proposition

Let $\mathcal{Q}$ be an n-tuple of points in $\mathbb{P}^r$ over $\mathbb{F}$ not in a hyperplane
Then the complexity of the computation of $I_2(\mathcal{Q})$ is at most $\mathcal{O}(n^4)$

Reverse engineering AG codes - 5

C is called very strong algebraic-geometric (VSAG)

if $C = C_L(\mathcal{X}, \mathcal{P}, E)$ and the curve $\mathcal{X}$ has genus g
$\mathcal{P}$ consists of n points and E has degree m such that

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

The dual of a VSAG code is again VSAG

Reverse engineering AG codes - 6

Main Theorem

Let C be a VSAG code

Then a VSAG representation of C can be obtained efficiently from its generator matrix

Moreover all VSAG representations of C are strict isomorphic

Error-correcting pair from VSAG code - 1

Shortcut via t-ECP pair (A, B) in $\mathbb{F}_q^n$
Bypassing triple $(\mathcal{X}, \mathcal{P}, E)$ and Riemann–Roch spaces

	$\mathbb{F}_q(\mathcal{X})$	$\mathbb{F}_q^n$
		$C = C_L(\mathcal{X}, \mathcal{P}, E)^{\perp}$
$(\mathcal{X}, \mathcal{P}, E)$	$L(E)$	$C_L(\mathcal{X}, \mathcal{P}, E)$
$(\mathcal{X}, \mathcal{P}, E-(t+g)P_1)$	$L(E-(t+g)P_1)$	$A = C_L(\mathcal{X}, \mathcal{P}, E-(t+g)P_1)$
$(\mathcal{X}, \mathcal{P}, (t+g)P_1)$	$L((t+g)P_1)$	$B = C_L(\mathcal{X}, \mathcal{P}, (t+g)P_1)$

Error-correcting pair from VSAG code - 2

In fact, A is the space of those code words in $C^{\perp}$
that are zero at the first position with multiplicity $m - t - g$
This multiplicity can be controlled
since we computed $I_2(\mathcal{Q})$ efficiently

Define $B_0 = \langle A * C \rangle^{\perp}$
Then $B_0^{\perp} = \langle A * C \rangle \subseteq B^{\perp}$
So $d(B_0^{\perp}) \geq d(B^{\perp}) > t$
Hence (A, B_0) is a t-ECP for C

Similarly decode up to $\lfloor (d^* - 1)/2 \rfloor$ errors with majority coset decoding

Questions?