How arithmetic and geometry make error correcting codes better?

Alain Couvreur

Conference: Curves Over Finite Fields : Past, Present and Future
Celerating the publication of Serre’s book on the topic

May 24–25, 2021
1. Coding theory

2. Algebraic geometry enters the game

3. New issues on AG codes
   - Component wise products of codes
   - Algebraic geometry codes and public key cryptography
   - Algebraic geometry codes for distributed storage
1 Coding theory

2 Algebraic geometry enters the game

3 New issues on AG codes
   - Component wise products of codes
   - Algebraic geometry codes and public key cryptography
   - Algebraic geometry codes for distributed storage
It started with…

Claude Shannon

Richard Hamming
Codes have been useful for...

**Wireless communication**

Mariner 9, used a [32, 6, 16] Reed–Muller code

**Telecommunication**

French Minitel, used a [127, 120, 3] BCH code

**Storage**

CD’s used, Reed–Solomon codes over $\mathbb{F}_{256}$
They are also useful for

- Secret key cryptography;
- Post quantum public key cryptography;
- Cloud storage;
- Secure multiparty computation;
- etc.
Coding theory is also...

... a rich theory with interactions with

- information theory;
- combinatorics;
- number theory;
- algebraic geometry.
By the way, what is a code?

**Definition 1**

A (linear) code is a vector subspace $\mathcal{C} \subseteq \mathbb{F}_q^n$.

- the integer $n$ is called its length;
- its dimension is denoted by $k$;
- its minimum distance $d$ is defined as

$$d \overset{\text{def}}{=} \min_{x \neq y \in \mathcal{C}} \#\{i \mid x_i \neq y_i\} = \min_{c \in \mathcal{C} \neq \{0\}} \#\{i \mid c_i \neq 0\}.$$

This triple of parameters is usually denoted as $[n, k, d]_q$. 
Algebraic construction — Reed–Solomon codes (1960)

Optimal parameters (MDS code);
Benefit from efficient decoding algorithms (Berlekamp 1968);
Drawback: require $n \leq q$. 

$x_1 \quad x_2 \quad x_3 \quad \ldots \quad x_n$
Algebraic construction — Reed–Solomon codes (1960)

\[ f \in F_q[X]^<k \]

Optimal parameters (MDS code);
Benefit from efficient decoding algorithms (Berlekamp 1968);
Drawback: require \( n \leq q \).

\[ x_1 \quad x_2 \quad x_3 \quad \ldots \quad x_n \]
Algebraic construction — Reed–Solomon codes (1960)

$f \in \mathbb{F}_q[X]_{<k}$

- Optimal parameters (MDS code);
- Benefit from efficient decoding algorithms (Berlekamp 1968);
- Drawback: \( n \leq q \).
Algebraic construction — Reed–Solomon codes (1960)

\[
\begin{align*}
&\text{RS}_k(x) \overset{\text{def}}{=} \{(f(x_1), f(x_2), f(x_3), \ldots, f(x_n)) \mid f \in \mathbb{F}_q[X]_{<k}\} \\
&f \in \mathbb{F}_q[X]_{<k} \\
x_1 & \quad x_2 & \quad x_3 & \quad \ldots & \quad x_n
\end{align*}
\]

Optimal parameters (MDS code); Benefit from efficient decoding algorithms (Berlekamp 1968); Drawback: require \(n \leq q\).
Algebraic construction — Reed–Solomon codes (1960)

\[ f \in \mathbb{F}_q[X]^{<k} \]

\[ \text{RS}_k(x) \overset{\text{def}}{=} \{(f(x_1), f(x_2), f(x_3), \ldots, f(x_n)) \mid f \in \mathbb{F}_q[X]^{<k}\} \]

- Optimal parameters (MDS code);
Algebraic construction — Reed–Solomon codes (1960)

\[ f \in \mathbb{F}_q[X]_k \]

\[ \text{RS}_k(x) \overset{\text{def}}{=} \{(f(x_1), f(x_2), f(x_3), \ldots, f(x_n)) \mid f \in \mathbb{F}_q[X]_k\} \]

- Optimal parameters (MDS code);
- Benefit from efficient decoding algorithms (Berlekamp 1968);
Algebraic construction — Reed–Solomon codes (1960)

$$f \in \mathbb{F}_q[X]_{<k}$$

$$\mathbf{RS}_k(x) \overset{\text{def}}{=} \{(f(x_1), f(x_2), f(x_3), \ldots, f(x_n)) \mid f \in \mathbb{F}_q[X]_{<k}\}$$

- Optimal parameters (MDS code);
- Benefit from efficient decoding algorithms (Berlekamp 1968);
- **Drawback:** require $n \leq q$. 
Coding theory

Algebraic geometry enters the game

New issues on AG codes
- Component wise products of codes
- Algebraic geometry codes and public key cryptography
- Algebraic geometry codes for distributed storage
Algebraic geometry codes (Goppa, 1981)

\[ f \in \mathbb{F}_q[X] \]

RS\(_k(x) \) def = \{(f(x_1), f(x_2), f(x_3), \ldots, f(x_n)) \mid f \in \mathbb{F}_q[X]_{<k}\}

Proposition 1

The parameters \([n, k, d]\) of this code satisfy:

\[ k \geq \deg G + 1 - g \]
\[ d \geq n - \deg G. \]
Algebraic geometry codes (Goppa, 1981)

\[ f \in L(G) \]

\[ C((P_i), G) \overset{\text{def}}{=} \{(f(P_1), f(P_2), f(P_3), \ldots, f(P_n)) \mid f \in L(G)\} \]
$f \in L(G)$

$C((P_i)_i, G) \overset{\text{def}}{=} \{(f(P_1), f(P_2), f(P_3), \ldots, f(P_n)) \mid f \in L(G)\}$

**Proposition 1**

The parameters $[n, k, d]$ of this code satisfy:

$$k \geq \deg G + 1 - g \quad d \geq n - \deg G.$$
Any code satisfies Singleton bound:

\[ n + 1 \geq k + d \]

AG codes satisfy

\[ k + d \geq n + 1 - g. \]
Any code satisfies Singleton bound:

\[ n + 1 \geq k + d \]

AG codes satisfy

\[ k + d \geq n + 1 - g. \]

“AG codes lie at distance \( g \) from optimum”.
Asymptotic parameters

**Proposition 2 (Gilbert Varshamov bound)**

For any $\varepsilon > 0$ and any $0 \leq \delta \leq 1 - \frac{1}{q}$, there exists a sequence of codes $(C_s)_s$ with

$$n_s \to +\infty, \quad \frac{d_s}{n_s} \to \delta, \quad \text{and} \quad \frac{k_s}{n_s} \to R \geq 1 - H_q(\delta) - \varepsilon.$$ 

Moreover, for a random choice of $C_s$ of rate $1 - H_q(\delta)$,

$$\text{Prob}\left[ n_s(\delta - \varepsilon) \leq d_s \leq n_s(\delta + \varepsilon) \right] \to 1$$
The typical behaviour of codes
With AG codes

AG codes satisfy

\[ k + d \geq n + 1 - g. \]

Denoting by \( A(q) \) the Ihara constant:

\[
A(q) \overset{\text{def}}{=} \lim_{g \to +\infty} \max_{\text{Genus}(X) = g} \frac{\# X(\mathbb{F}_q)}{g},
\]

we prove the existence of sequences of codes satisfying

\[
R + \delta \geq 1 - \frac{1}{A(q)}.
\]
With AG codes

\[ R + \delta \geq 1 - \frac{1}{A(q)}. \]

By Drinfeld–Vlăduţ bound, \( A(q) \leq \sqrt{q} - 1 \). Moreover:


*For \( q \) a square, \( A(q) = \sqrt{q} - 1 \).*
The famous picture (for $q = 49$)
How to get families of curves with many points?

  - $X_0(\ell)$ for $\ell \to +\infty$ yield $A(p^2) = p - 1$ for $p$ prime;
  - Shimura curves extend to $A(q)$ for arbitrary prime power $q$

- Recursive towers (Garcia, Stichtenoth 1995);

- Class field theory (Serre).

$$A(q) \geq c \cdot \log(q) \quad \text{with} \quad c = \frac{1}{96}.$$
Why only curves?

- The construction extends straightforwardly to higher dimensional varieties (Manin 1984)
- Even for surfaces, the estimate of the minimum distance gets very hard.

Codes from curves:
Why only curves?

- The construction extends straightforwardly to higher dimensional varieties (Manin 1984)
- Even for surfaces, the estimate of the minimum distance gets very hard.

Codes from curves:

\[ \text{Zéros de } f \in L(G) \]
Why only curves?

- The construction extends straightforwardly to higher dimensional varieties (Manin 1984).
- Even for surfaces, the estimate of the minimum distance gets very hard.

Codes from curves:

\[ \text{Zéros de } f \in L(G) \]
Why only curves?

- The construction extends straightforwardly to higher dimensional varieties (Manin 1984)
- Even for surfaces, the estimate of the minimum distance gets very hard.

Codes from surfaces:
Why only curves?

- The construction extends straightforwardly to higher dimensional varieties (Manin 1984)
- Even for surfaces, the estimate of the minimum distance gets very hard.

Codes from surfaces:

\[ Zéros de f \in L(G) \]
Why only curves?

- The construction extends straightforwardly to higher dimensional varieties (Manin 1984)
- Even for surfaces, the estimate of the minimum distance gets very hard.

Codes from surfaces:
Why only curves?

- The construction extends straightforwardly to higher dimensional varieties (Manin 1984)
- Even for surfaces, the estimate of the minimum distance gets very hard.

Codes from surfaces:

\[ \text{Zéros de } f \in L(G) \]
Why only curves?

- The construction extends straightforwardly to higher dimensional varieties (Manin 1984)
- Even for surfaces, the estimate of the minimum distance gets very hard.

Codes from surfaces:

Zéros de $f \in L(G)$
Why only curves?

- The construction extends straightforwardly to higher dimensional varieties (Manin 1984)
- Even for surfaces, the estimate of the minimum distance gets very hard.

Codes from surfaces:
Why only curves?

- The construction extends straightforwardly to higher dimensional varieties (Manin 1984)
- Even for surfaces, the estimate of the minimum distance gets very hard.
Why only curves?

- The construction extends straightforwardly to higher dimensional varieties (Manin 1984)
- Even for surfaces, the estimate of the minimum distance gets very hard.

Codes from surfaces:

\[ \text{Zéros de } f \in L(G) \]

\[ \rightarrow \text{Need for estimating: } \max_{D \in |G|} \# \text{Supp}(D)(\mathbb{F}_q). \]
From surfaces Need for estimating the maximal number of rational points of a curve in a linear system:

- Need for counting the maximum number of irreducible components (related to the Neron Severi group of the surface (Voloch & Zarzar 2009));
- Then apply Weil and Serre bound to the components.

\[ \#X(\mathbb{F}_q) \leq q + 1 + g \lfloor 2\sqrt{q} \rfloor \]

- or use other estimates: Stöhr–Voloch, Homma–Kim.
Back to codes from curves
Back to curves, what about decoding?

**Problem.** Let $y = c + e$ with $c \in C(\mathcal{P}, G)$ and $w(e) = t$. How to find $c$?

**Idea.** Fix an extra divisor $F$ and compute $Q_0 \in L(F + G)$ and $Q_1 \in L(F)$ such that

$$\forall i \in \{1, \ldots, n\}, \quad Q_0(P_i) + Q_1(P_i)y_i = 0$$

**Theorem 2**

Let $f \in L(G)$ such that $c = (f(P_1), \ldots, f(P_n))$. If $t \leq \frac{n - \deg G}{2} - \frac{g}{2}$ and $\deg F \geq t + g$ then

$$Q_0 + Q_1 f = 0$$
History of decoding

- Up to $\frac{n-\deg G}{2} - \frac{g}{2}$:

- Up to $\frac{n-\deg G}{2}$:

- Beyond half the minimum distance (list decoding):
  - Sudan 1997 (only Reed–Solomon codes);
  - Shokrollahi Wasserman 1999;
  - Guruswami Sudan 1999.
1 Coding theory

2 Algebraic geometry enters the game

3 New issues on AG codes
   - Component wise products of codes
   - Algebraic geometry codes and public key cryptography
   - Algebraic geometry codes for distributed storage
1 Coding theory

2 Algebraic geometry enters the game

3 New issues on AG codes
   - Component wise products of codes
   - Algebraic geometry codes and public key cryptography
   - Algebraic geometry codes for distributed storage
Another point of view for decoding

Data: \( y = c + e \) with \( c = (f(P_1), \ldots, f(P_n)) \) and \( f \in L(G) \) and \( w(e) = t \).

- **Non linear problem:** find \( Q_1 \in L(F) \) and \( f \in L(G) \) such that
  \[ \forall i \in \{1, \ldots, n\}, \; Q_1(P_i)y_i = Q_1(P_i)f(P_i). \]

  i.e. one looks for \( f \) and for a function \( Q \in L(F) \) vanishing at the error positions.

- **Linearisation:** Set \( Q_0 \stackrel{\text{def}}{=} Q_1f \) an look for the pair \((Q_0, Q_1)\).
Another point of view for decoding

Data: $y = c + e$ with $c = (f(P_1), \ldots, f(P_n))$ and $f \in L(G)$ and $w(e) = t$.

- **Non linear problem:** find $Q_1 \in L(F)$ and $f \in L(G)$ such that

  $$\forall i \in \{1, \ldots, n\}, \ Q_1(P_i)y_i = Q_1(P_i)f(P_i).$$

  *i.e. one looks for $f$ and for a function $Q \in L(F)$ vanishing at the error positions.*

- **Linearisation:** Set $Q_0 \overset{\text{def}}{=} Q_1 f$ an look for the pair $(Q_0, Q_1)$.

Linearisation is efficient since

$$Q_0 \in \text{Span}\{uv \mid u \in L(F), \ v \in L(G)\} \subseteq L(F + G).$$
Another point of view of decoding

**Definition 2**

Let $\mathcal{C}, \mathcal{D} \subseteq \mathbb{F}_q^n$ be two codes, we denote by $\mathcal{C} \star \mathcal{D}$ the code

$$\mathcal{C} \star \mathcal{D} \overset{\text{def}}{=} \text{Span} \{ (c_1 d_1, \ldots, c_n, d_n) \mid c \in \mathcal{C}, \ d \in \mathcal{D} \}$$

**In summary**

- Codes from curves can be decoded because $\star$–products of such codes are “small”.
- In 1992, Koetter and Pellikaan proposed an abstract version (Error correcting pairs) of AG decoding only requiring such a behaviour w.r.t the $\star$–product.
- Bad news for codes from higher dimensional varieties.
Similarly to additive combinatorics, there is a Kneser theorem for star–products:

**Theorem 3**

Let \( C, D \subseteq F^*_q \), then

\[
\dim C \star D \geq \dim C + \dim D - \dim \{ x \in F^*_q \mid x \star C \star D \subseteq C \star D \}
\]

**Question:** Which codes satisfy \( \dim C \star C \leq 2 \dim C - 1 + \gamma \) with \( \gamma \) “small”?

**Same question on the level of function fields:** Given a function field \( F \) over \( F_q \), which finite subspaces \( S \subseteq F \) satisfy \( \dim S \cdot S \leq 2 \dim S - 1 + \gamma \)?
Theorem 4 (Bachoc, C., Zémor, 2019)

Let $k$ be a perfect field and $F$ be a finitely generated extension of $k$ with positive transcendence degree. Let $S \subseteq F$ be a finite dimensional subspace such that $1 \in S$ and $k(S) = F$. If

$$\dim S \cdot S \leq 2 \dim S - 1 + \gamma$$

and $\gamma \leq \dim S - 3$, then

- $F/k$ has transcendence degree 1;
- moreover, if $\gamma = 1$, then $F$ has genus 0 or 1.

Open question: what about higher $\gamma$'s? what about codes with a small $C^\ast C$? do they all come from curves?
A partial answer

**Theorem 4 (Bachoc, C., Zémor, 2019)**

Let $k$ be a perfect field and $F$ be a finitely generated extension of $k$ with positive transcendence degree. Let $S \subseteq F$ be a finite dimensional subspace such that $1 \in S$ and $k(S) = F$. If

$$\dim S \cdot S \leq 2 \dim S - 1 + \gamma$$

and $\gamma \leq \dim S - 3$, then

- $F/k$ has transcendence degree 1;
- moreover, if $\gamma = 1$, then $F$ has genus 0 or 1.

**Open question:**

- what about higher $\gamma$'s?
- what about codes with a small $\mathcal{C} \star \mathcal{C}$? do they all come from curves?
On the level of codes

Given a generator matrix

\[
G = \begin{pmatrix}
g_{11} & \cdots & g_{n1} \\
\vdots & & \vdots \\
g_{k1} & \cdots & g_{kn}
\end{pmatrix} = \begin{pmatrix}
f_1(P_1) & \cdots & f_1(P_n) \\
\vdots & & \vdots \\
f_k(P_1) & \cdots & f_k(P_n)
\end{pmatrix}
\]

- The columns may be regarded as points of \( \mathbb{P}^{k-1} \).
On the level of codes

Given a generator matrix

\[
G = \begin{pmatrix}
g_{11} & \cdots & g_{n1} \\
\vdots & & \vdots \\
g_{k1} & \cdots & g_{kn}
\end{pmatrix} = \begin{pmatrix}
f_1(P_1) & \cdots & f_1(P_n) \\
\vdots & & \vdots \\
f_k(P_1) & \cdots & f_k(P_n)
\end{pmatrix}
\]

- The columns may be regarded as points of \( \mathbb{P}^{k-1} \).
- They are the images of the \( P_i \) by the “embedding” given by \( G: X \to \mathbb{P}(H^0(X, G)) \).
On the level of codes

Given a generator matrix

\[
G = \begin{pmatrix}
g_{11} & \cdots & g_{n1} \\
\vdots & \ddots & \vdots \\
g_{k1} & \cdots & g_{kn}
\end{pmatrix} = \begin{pmatrix}
f_1(P_1) & \cdots & f_1(P_n) \\
\vdots & \ddots & \vdots \\
f_k(P_1) & \cdots & f_k(P_n)
\end{pmatrix}
\]

- The columns may be regarded as points of \(\mathbb{P}^{k-1}\).
- They are the images of the \(P_i\) by the “embedding” given by \(G: X \rightarrow \mathbb{P}(H^0(X, G))\).
- Consider the AG code \(\mathcal{C}\) and the exact sequence

\[
0 \rightarrow I_2(\mathcal{C}) \rightarrow S^2\mathcal{C} \rightarrow \mathcal{C} \ast \mathcal{C} \rightarrow 0
\]

\(V(I_2(\mathcal{C}))\) contains the curve \(X\). There is equality if \(2g + 1 \leq \text{deg } G < \frac{n}{2}\) (Márquez–Corbella, Martínez–Moro, Pellikaan, 2013).
On the level of codes

Given a generator matrix

\[
G = \begin{pmatrix}
g_{11} & \cdots & g_{n1} \\
\vdots & \ddots & \vdots \\
g_{k1} & \cdots & g_{kn}
\end{pmatrix} = \begin{pmatrix}
f_1(P_1) & \cdots & f_1(P_n) \\
\vdots & \ddots & \vdots \\
f_k(P_1) & \cdots & f_k(P_n)
\end{pmatrix}
\]

- The columns may be regarded as points of \( \mathbb{P}^{k-1} \).
- They are the images of the \( P_i \) by the “embedding” given by \( G : X \rightarrow \mathbb{P}(H^0(X, G)) \).
- Consider the AG code \( \mathcal{C} \) and the exact sequence

\[
0 \rightarrow I_2(\mathcal{C}) \rightarrow S^2\mathcal{C} \rightarrow \mathcal{C} \star \mathcal{C} \rightarrow 0
\]

\( V(I_2(\mathcal{C})) \) contains the curve \( X \). There is equality if \( 2g + 1 \leq \deg G < \frac{n}{2} \) (Márquez–Corbella, Martínez–Moro, Pellikaan, 2013).
- What about \( V(I_2(\mathcal{C})) \) for codes with a “small square” and which do not \( \text{à priori} \) come from a curve?
On the level of codes

What about $V(l_2(C))$ for codes with a “small square” and which do not à priori come from a curve?
On the level of codes

- What about $V(I_2(C))$ for codes with a “small square” and which do not à priori come from a curve?
- **Observation:** if $V(I_2(C))$ is irreducible, then

$$L(G) \cong H^0(\mathbb{P}^{k-1}, \mathcal{O}(1)) \rightarrow C$$

$$H^0(V(I_2(C)), \mathcal{O}(1))$$
On the level of codes

- What about $V(I_2(C))$ for codes with a “small square” and which do not à priori come from a curve?
- **Observation:** if $V(I_2(C))$ is irreducible, then

\[ L(G) \simeq H^0(\mathbb{P}^{k-1}, O(1)) \to C \]

\[ H^0(V(I_2(C)), O(1)) \]

$L(G)$ identifies to a subspace with a small square generating the function field of $V(I_2(C))$. $\implies V(I_2(C))$ is a curve (using Bachoc, C., Zémor).
On the level of codes

The condition of irreducibility is fundamental:

- Let $V \subseteq \mathbb{P}^3$ be the union of a plane and a line not contained in it;

We have $\dim C = 4$ and $\dim C^* = 7$. And $V(I_2(C)) = V$ is reducible!
On the level of codes

The condition of irreducibility is fundamental:

- Let $V \subseteq \mathbb{P}^3$ be the union of a plane and a line not contained in it;
- Let $\mathcal{C}$ be the code obtained by evaluating $H^0(\mathbb{P}^3, \mathcal{O}(1))$ at the points of $V$. 
On the level of codes

The condition of irreducibility is fundamental:

- Let \( V \subseteq \mathbb{P}^3 \) be the union of a plane and a line not contained in it;
- Let \( C \) be the code obtained by evaluating \( H^0(\mathbb{P}^3, \mathcal{O}(1)) \) at the points of \( V \).
- We have \( \dim C = 4 \) and \( \dim C \star C = 7 \).
On the level of codes

The condition of irreducibility is fundamental:

- Let \( V \subseteq \mathbb{P}^3 \) be the union of a plane and a line not contained in it;
- Let \( C \) be the code obtained by evaluating \( H^0(\mathbb{P}^3, \mathcal{O}(1)) \) at the points of \( V \).
- We have \( \dim C = 4 \) and \( \dim C \star C = 7 \).
- And \( V(I_2(C)) \neq V \) is reducible!
In summary

- How to classify codes $\mathcal{C}$ such that $\dim \mathcal{C} \star \mathcal{C} = 2 \dim \mathcal{C} - 1 + \gamma$ with a “small” $\gamma$?
- Under which additional hypotheses may we expect an irreducible $V(I_2(\mathcal{C}))$?
1 Coding theory

2 Algebraic geometry enters the game

3 New issues on AG codes
   - Component wise products of codes
   - Algebraic geometry codes and public key cryptography
   - Algebraic geometry codes for distributed storage
McEliece encryption scheme

- Dates back from 1978 (McEliece);

- Security relies on:
  - the computational hardness of decoding a random code;
  - the computational hardness of distinguishing some structured codes (which we can decode) with random codes.

Properties

Advantages
- Post quantum;
- Fast.

Drawbacks
- Requires huge key sizes.

Janwa, Moreno (1996) suggest to use AG codes.
McEliece encryption scheme

- Dates back from 1978 (McEliece);
- Security relies on
  - the computational hardness of decoding a random code;
  - the computational hardness of distinguishing some structured codes (which we can decode) with random codes.

Properties

Advantages

- Post quantum;
- Fast.

Drawbacks

- Requires huge key sizes.

Janwa, Moreno (1996) suggest to use AG codes.
McEliece encryption scheme

- Dates back from 1978 (McEliece);
- Security relies on
  - the computational hardness of decoding a random code;
  - the computational hardness of distinguishing some structured codes (which we can decode) with random codes.

Properties

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Drawbacks</th>
</tr>
</thead>
<tbody>
<tr>
<td>▶ Post quantum;</td>
<td>▶ Requires huge key sizes.</td>
</tr>
<tr>
<td>▶ Fast.</td>
<td></td>
</tr>
</tbody>
</table>
McEliece encryption scheme

- Dates back from 1978 (McEliece);
- Security relies on
  - the computational hardness of decoding a random code;
  - the computational hardness of distinguishing some structured codes (which we can decode) with random codes.

Properties

**Advantages**

- Post quantum;
- Fast.

**Drawbacks**

- Requires huge key sizes.

- Janwa, Moreno (1996) suggest to use AG codes.
Janwa, Moreno (1996) suggest to use AG codes, in 3 different manners.

- Raw AG codes;
- Concatenated AG codes.
- $C \cap \mathbb{F}_q^n$ with $C$ an AG code over $\mathbb{F}_{q^m}$ (direct generalisation of McEliece);
Janwa, Moreno (1996) suggest to use AG codes, in 3 different manners.

- Raw AG codes; [UNSECURE]
- Concatenated AG codes. [UNSECURE] (Sendrier, 1994)
- $\mathcal{C} \cap \mathbb{F}_q^n$ with $\mathcal{C}$ an AG code over $\mathbb{F}_{q^m}$ (direct generalisation of McEliece);
Janwa, Moreno (1996) suggest to use AG codes, in 3 different manners.

- Raw AG codes; [UNSECURE]
- Concatenated AG codes. [UNSECURE] (Sendrier, 1994)
- $\mathcal{C} \cap \mathbb{F}_q^n$ with $\mathcal{C}$ an AG code over $\mathbb{F}_{q^m}$ (direct generalisation of McEliece);

Attacks on raw AG codes:

- $g = 0$: Sidelnikov Shestakov 1992;
- $g = 1, 2$: Faure, Minder, 2009;
- Any $g$: C., Márquez–Corbella, Pellikaan 2017.
Janwa, Moreno (1996) suggest to use AG codes, in 3 different manners.

- Raw AG codes; [UNSECURE]
- Concatenated AG codes. [UNSECURE] (Sendrier, 1994)
- $\mathcal{C} \cap \mathbb{F}_q^n$ with $\mathcal{C}$ an AG code over $\mathbb{F}_{q^m}$ (direct generalisation of McEliece);

Attacks on raw AG codes:

- $g = 0$: Sidelnikov Shestakov 1992;
- $g = 1, 2$: Faure, Minder, 2009;
- Any $g$: C., Márquez–Corbella, Pellikaan 2017.

Related to the fact that $\mathcal{C} \ast \mathcal{C}$ behaves very differently for AG codes and random codes.
It remains the case of *subfield subcodes* : $\mathcal{C} \cap \mathbb{F}^n_q$ for $\mathcal{C}$ an AG code over $\mathbb{F}_{q^m}$.

- Which $m$ should be chosen? For $g = 0$, $m = 2$ has been proved to be weak.
- How to reduce key sizes? Use curves with automorphisms?
Coding theory

Algebraic geometry enters the game

New issues on AG codes
- Component wise products of codes
- Algebraic geometry codes and public key cryptography
- Algebraic geometry codes for distributed storage
Cloud storage
Data is distributed over various servers;
Cloud storage

- Data is distributed over various servers;
- In case of a server failure, data should be reconstructed while limiting the bandwidth.
Cloud storage

- Data is distributed over various servers;
- In case of a server failure, data should be reconstructed while limiting the bandwidth.
- **Model** each server carries one symbol of a codeword.
  - How to reconstruct a symbol from a limited amount of other symbols?
Data is distributed over various servers;
In case of a server failure, data should be reconstructed while limiting the bandwidth.
**Model** each server carries one symbol of a codeword.
  ▶ How to reconstruct a symbol from a limited amount of other symbols?
With Reed–Solomon codes (AG codes with $g = 0$) reconstructing one symbol requires the knowledge of $k$ symbols (cannot be worse).
Definition 3

A locally recoverable code (LRC) with locality $\ell$ is a code $C \subseteq \mathbb{F}_q^n$ such that for any $i \in \{1, \ldots, n\}$ there is at least one subset $A(i) \subseteq \{1, \ldots, n\}$ containing $i$ such that $\#A(i) \leq \ell + 1$ and $C|_{A(i)}$ has minimum distance $\geq 2$. The set $A(i)$ is called a recovery set for $i$. 

Theorem 1 (Singleton bound for LRC's)

Let $C \subseteq \mathbb{F}_q^n$ be an $[n, k, d]$ LRC with locality $\ell$, then $d \leq n - k - \lceil k \ell \rceil + 2$. 

A. Couvreur

May 24–25, 2021 41 / 46
Definition 3

A locally recoverable code (LRC) with locality $\ell$ is a code $C \subseteq \mathbb{F}_q^n$ such that for any $i \in \{1, \ldots, n\}$ there is at least one subset $A(i) \subseteq \{1, \ldots, n\}$ containing $i$ such that $\#A(i) \leq \ell + 1$ and $C|_{A(i)}$ has minimum distance $\geq 2$. The set $A(i)$ is called a recovery set for $i$.

Theorem 1 (Singleton bound for LRC’s)

Let $C \subseteq \mathbb{F}_q^n$ be an $[n, k, d]$ LRC with locality $\ell$, then

$$d \leq n - k - \left\lceil \frac{k}{\ell} \right\rceil + 2$$
Consider

- an automorphism $\sigma : \mathbb{P}^1 \to \mathbb{P}^1$ of order $\ell + 1$ (e.g. $z \mapsto \zeta z$ with $\zeta^{\ell+1} = 1$);
Consider

- an automorphism $\sigma : \mathbb{P}^1 \to \mathbb{P}^1$ of order $\ell + 1$ (e.g. $z \mapsto \zeta z$ with $\zeta^{\ell+1} = 1)$;
- A $\sigma$–invariant polynomial $g \in \mathbb{F}_q[X]$ of degree $\ell + 1$ (e.g. $g(X) = X^{\ell+1}$).
Tamo–Barg construction

Consider

- an automorphism $\sigma : \mathbb{P}^1 \to \mathbb{P}^1$ of order $\ell + 1$ (e.g. $z \mapsto \zeta z$ with $\zeta^{\ell+1} = 1$);
- A $\sigma$–invariant polynomial $g \in \mathbb{F}_q[X]$ of degree $\ell + 1$ (e.g. $g(X) = X^{\ell+1}$).
- Rational points $P_{11}, \ldots, P_{\frac{n}{\ell+1}, \ell+1}$ gathered as $\sigma$–orbits.
Consider

- an automorphism $\sigma : \mathbb{P}^1 \rightarrow \mathbb{P}^1$ of order $\ell + 1$ (e.g. $z \mapsto \zeta z$ with $\zeta^{\ell+1} = 1$);
- A $\sigma$–invariant polynomial $g \in \mathbb{F}_q[X]$ of degree $\ell + 1$ (e.g. $g(X) = X^{\ell+1}$).
- Rational points $P_{11}, \ldots, P_{n^{\ell+1},\ell+1}$ gathered as $\sigma$–orbits.
- Define the code

$$
\mathcal{C} \overset{\text{def}}{=} \left\{ (f(P_{11}), \ldots, f(P_{n^{\ell+1},\ell+1})) \mid f(X) = \sum_{i=0}^{\ell-1} \sum_{j=0}^{k-1} a_{ij} X^i g(X)^j \right\}.
$$
Consider

- an automorphism $\sigma : \mathbb{P}^1 \to \mathbb{P}^1$ of order $\ell + 1$ (e.g. $z \mapsto \zeta z$ with $\zeta^{\ell + 1} = 1$);
- A $\sigma$–invariant polynomial $g \in \mathbb{F}_q[X]$ of degree $\ell + 1$ (e.g. $g(X) = X^{\ell + 1}$).
- Rational points $P_{11}, \ldots, P_{\frac{n}{\ell + 1}, \ell + 1}$ gathered as $\sigma$–orbits.
- Define the code

$$C \overset{\text{def}}{=} \left\{ (f(P_{11}), \ldots, f(P_{\frac{n}{\ell + 1}, \ell + 1})) \mid f(X) = \sum_{i=0}^{\ell - 1} \sum_{j=0}^{\frac{k - 1}{\ell}} a_{ij} X^i g(X)^j \right\}.$$ 

- **Idea:** Restrict to an orbit $A$, where $g|_A = c$, we get $f|_A = \sum_{i,j} a_{ij} c^j X^i$. 

\[TAMO–BARG\text{ construction}\]
Consider

- an automorphism $\sigma : \mathbb{P}^1 \to \mathbb{P}^1$ of order $\ell + 1$ (e.g. $z \mapsto \zeta z$ with $\zeta^{\ell+1} = 1$);
- A $\sigma$–invariant polynomial $g \in \mathbb{F}_q[X]$ of degree $\ell + 1$ (e.g. $g(X) = X^{\ell+1}$).
- Rational points $P_{11}, \ldots, P_{\frac{n}{\ell+1},\ell+1}$ gathered as $\sigma$–orbits.
- Define the code

$$C \overset{\text{def}}{=} \left\{ (f(P_{11}), \ldots, f(P_{\frac{n}{\ell+1},\ell+1})) \mid f(X) = \sum_{i=0}^{\ell-1} \sum_{j=0}^{k-1} a_{ij} X^i g(X)^j \right\}.$$

- **Idea:** Restrict to an orbit $A$, where $g|_A = c$, we get $f|_A = \sum_{i,j} a_{ij} c^j X^i$.
  - The degree collapses!
Consider

- an automorphism $\sigma : \mathbb{P}^1 \to \mathbb{P}^1$ of order $\ell + 1$ (e.g. $z \mapsto \zeta z$ with $\zeta^{\ell+1} = 1$);
- A $\sigma$–invariant polynomial $g \in \mathbb{F}_q[X]$ of degree $\ell + 1$ (e.g. $g(X) = X^{\ell+1}$).
- Rational points $P_{11}, \ldots, P_{\frac{n}{\ell+1}, \ell+1}$ gathered as $\sigma$–orbits.
- Define the code

$$C = \left\{ (f(P_{11}), \ldots, f(P_{\frac{n}{\ell+1}, \ell+1})) \mid f(X) = \sum_{i=0}^{\ell-1} \sum_{j=0}^{k-1} a_{ij} X^i g(X)^j \right\}.$$ 

**Idea:** Restrict to an orbit $A$, where $g|_A = c$, we get $f|_A = \sum_{i,j} a_{ij} c^j X^i$.

- The degree collapses!
- Optimal construction for the “Singleton–like bound for LRC’s”.

▶ The degree collapses!
Consider a Galois cover of degree $\ell + 1$ (previous example, $X = Y = \mathbb{P}^1$):

$$
\begin{array}{c}
Y \\
\downarrow \phi \\
X
\end{array}
$$

Definition 4 (Barg, Tamo, Vlăduţ)

Let $P_{11}, \ldots, P_{1,\ell+1}, \ldots, P_{n,\ell+1} \in Y(F_q)$ be the fibres of totally splitting points of $X$. Let $G$ be a divisor of $X$ and $x \in F_q(Y)$ be a primitive element of $F_q(Y)/F_q(X)$. Define the code $C$ as

$$
C := \left\{ (f(P_{ij}))_{i,j} \mid f = \ell - 1 \sum_{i=0}^{\ell} (\phi^* f_i) \cdot x^i, f_i \in L(G) \right\}.
$$
Consider a Galois cover of degree $\ell + 1$ (previous example, $X = Y = \mathbb{P}^1$):

$$
\begin{array}{ccc}
Y_P & \longrightarrow & Y \\
\downarrow & & \downarrow \phi \\
P & \longrightarrow & X
\end{array}
$$

with functions whose restrictions on fibres have “low degree”.

Definition 4 (Barg, Tamo, Vlăduț)

Let $P_{11}, \ldots, P_{1,\ell+1}, \ldots, P_{n\ell+1}, \ldots, P_{n,\ell+1} \in Y(F_{q^1})$ be the fibres of totally splitting points of $X$. Let $G$ be a divisor of $X$ and $x \in F_{q^1}(Y)$ be a primitive element of $F_{q^1}(Y)/F_{q^1}(X)$. Define the code $C_{\text{def}} = \{ (f(P_{ij}))_{i,j} | f = \ell - 1 \sum_{i=0}^{\ell} (\phi^* f_i) \cdot x^i, f_i \in L(G) \}$.
Consider a Galois cover of degree \( \ell + 1 \) (previous example, \( X = Y = \mathbb{P}^1 \)):

\[
\begin{array}{ccc}
Y_P & \longrightarrow & Y \\
\downarrow & & \downarrow \phi \\
P & \longrightarrow & X
\end{array}
\]

with functions whose restrictions on fibres have “low degree”.

**Definition 4 (Barg, Tamo, Vlăduţ)**

Let \( P_{11}, \ldots, P_{1,\ell+1}, \ldots, P_{n,\ell+1} \in Y(\mathbb{F}_q) \) be the fibres of totally splitting points of \( X \). Let \( G \) be a divisor of \( X \) and \( x \in \mathbb{F}_q(Y) \) be a primitive element of \( \mathbb{F}_q(Y)/\mathbb{F}_q(X) \). Define the code

\[
\mathcal{C} \overset{\text{def}}{=} \left\{ (f(P_{ij}))_{i,j} \left| \begin{array}{c}
f = \sum_{i=0}^{\ell-1} (\phi^* f_i) \cdot x^i, \quad f_i \in L(G)
\end{array} \right. \right\}.
\]
What else?

- Extension to Galois covers of surfaces (Barg, Haymaker, Howe, Matthews, Várilly–Alvarado, 2017);

- Using a fibred surfaces point of view (Salgado, Várilly–Alvarado, Voloch, 2019);

- What about multiple recovery sets?

- Using fibre products of Galois covers to increase the number of recovery sets (Haymaker, Malmskøg, Matthews, 2018).
What else?

- Extension to Galois covers of surfaces (Barg, Haymaker, Howe, Matthews, Várilly–Alvarado, 2017);
- Using a fibred surfaces point of view (Salgado, Várilly–Alvarado, Voloch, 2019);
What else?

- Extension to Galois covers of surfaces (Barg, Haymaker, Howe, Matthews, Várilly-Alvarado, 2017);
- Using a fibred surfaces point of view (Salgado, Várilly-Alvarado, Voloch, 2019);

What about multiple recovery sets?
What else?

- Extension to Galois covers of surfaces (Barg, Haymaker, Howe, Matthews, Várilly–Alvarado, 2017);
- Using a fibred surfaces point of view (Salgado, Várilly–Alvarado, Voloch, 2019);

What about multiple recovery sets?
- Using fibre products of Galois covers to increase the number of recovery sets (Haymaker, Malmskøg, Matthews, 2018).
What else?

- Extension to Galois covers of surfaces (Barg, Haymaker, Howe, Matthews, Várilly–Alvarado, 2017);
- Using a fibred surfaces point of view (Salgado, Várilly–Alvarado, Voloch, 2019);

What about multiple recovery sets?
- Using fibre products of Galois covers to increase the number of recovery sets (Haymaker, Malmskjøg, Matthews, 2018).
Questions

- Is there a fully geometric interpretation of

\[ C \overset{\text{def}}{=} \left\{ (f(P_{ij}))_{i,j} \mid f = \sum_{i=0}^{\ell-1} (\phi^* f_i) \cdot x^i, \quad f_i \in L(G) \right\} \]?

Why shall we choose a primitive element?

- On the other hand is there a purely coding theoretic description of this construction? Fibre product of codes?

- What about non disjoint recovery sets? Using codes from surfaces with low \textit{intersectoral genus} (Little, Schenk, 2019)?
Thank you!