Algebraic manipulation detection and systems of sets in groups

Sophie Huczynska
University of St Andrews

Work described is joint with various combinations from:
Maura Paterson, Gary Mullen, Jim Davis, Chris Jefferson, Silvia Nepšinská and others...

March 2021

Information security

Information security is concerned with the safe and private transmission and storage of data.

Motivating questions include:

- How can a message be sent so that we can detect whether it has been changed during transmission?
- If we detect that a change has occurred, can we recover the original message - and if so, how?
- How can we encrypt messages/data so that they cannot feasibly be decrypted by anyone other than the intended recipient?
- . . . and many more.

Manipulation detection

In this talk, we consider an encoding system and how to design it to minimise the chances that an undetected change can occur.

Applies to various situations:

- message transmission which is subject to attack
- storage device which is subject to tampering

We will be thinking in terms of the message-sending scenario.

It is helpful to model the situation as a “game” between an encoder and an adversary who is trying to “cheat” the encoder.

Our focus is on algebraic manipulation detection (AMD) codes.

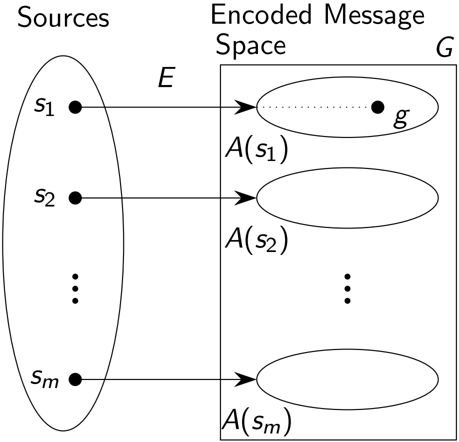
AMD code model

We will have:

- Set S of plaintext sources (the messages)
- Encoded message space G (finite group, written additively)
- Encoding function E (possibly randomized) maps source $s \in S$ to some $g \in G$
- For each source $s \in S$, subset $A(s)$ of G is the set of valid encodings of s
- Unique decodability: $A(s) \cap A(s') = \emptyset$ if $s \neq s'$, i.e. the sets of encodings do not overlap

Traditionally G abelian but the set-up is valid for non-abelian G.

Diagram



The "game"

AMD code

Adversary: chooses a value $\delta \in G \setminus \{0\}$ (their "manipulation")

Encoder: chooses source $s \in S$

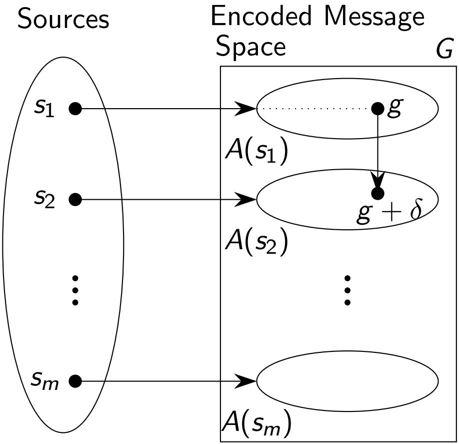
Encoder: s encoded by E to some $g \in A(s)$

Adversary: g is replaced by $g' = g + \delta$

Adversary wins if $g' \in A(s')$ for some $s' \neq s$

The adversary wins if they succeed in shifting the group element g into an element $g + \delta$ that's an encoding of a different source

Diagram



If message s_1 is sent and encoded as g, it will be incorrectly decoded to s_2 after this manipulation. In this case, adversary wins!

Model as system of sets

The AMD "game" can be modelled as a set-up in combinatorics.

We model the sender's choice of message probabilistically.

Let $\{A_1, \ldots, A_m\}$ be a disjoint collection of sets in G.

- Adversary chooses $\delta \in G \setminus \{0\}$
- Pick a set A_i uniformly at random (source)
- Then pick an element $d_i \in A_i$ uniformly at random (encoding)
- Adversary "wins" if $d_i + \delta \in A_j$ for some $j \neq i$

Adversary wins if δ occurs as a difference between our element in A_i and some element in A_j.

Need to look at the differences between elements of A_i and A_j.

Internal and external differences

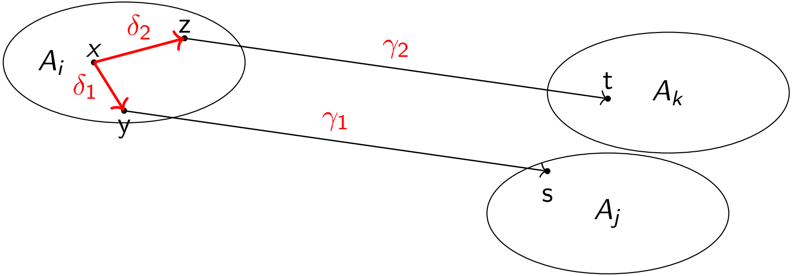
Suppose we have a disjoint family of subsets $A_1, \ldots, A_m$ of G

- For a fixed i, the differences between the elements of A_i are called internal differences:

$$I(A_i) := \{x - y : x, y \in A_i, x \neq y\}$$

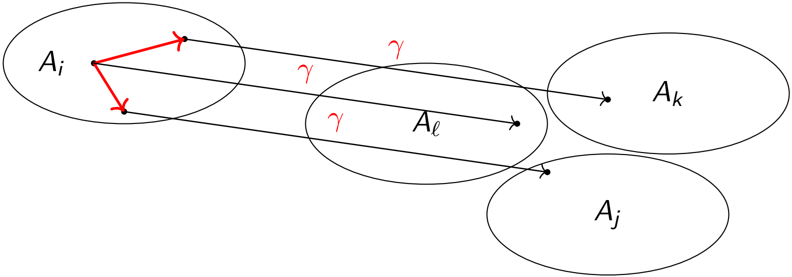
- For $i \neq j$, the differences between the elements of A_i and A_j are called external differences:

$$E(A_i, A_j) := \{x - y : x \in A_i, y \in A_j\}$$



In this diagram,

- δ_1 and δ_2 are internal differences in A_i
  $(x - y = \delta_1,\ x - z = \delta_2)$
- γ_1 and γ_2 are external differences out of A_i (to A_j, A_k resp.)
  $(y - s = \gamma_1,\ z - t = \gamma_2)$



For a disjoint family of sets $A_1, \ldots, A_m$, define the number of times a non-zero element γ occurs as an external difference out of A_i by

$$N_i(\gamma) := |\{(x, y) : x - y = \gamma, x \in A_i, y \in A_j, j \neq i\}|$$

In the example above, we show all occurrences of γ as an external difference out of A_i, so $N_i(\gamma) = 3$ here.

Success probability

Returning to our AMD code:

The probability that an adversary succeeds when they pick δ is

$$e_\delta = \frac{1}{m}\left(\frac{N_1(\delta)}{|A_1|} + \cdots + \frac{N_m(\delta)}{|A_m|}\right) \tag{1}$$

- Source i picked with probability $\frac{1}{m}$
- $N_i(\delta)$ of the possible $|A_i|$ encodings will lead to success for an adversary who picks δ

Which codes are optimal?

We seek AMD codes that are optimal (from sender's perspective).

We are considering an adversary who chooses δ uniformly at random (R-strategy)

Definition

An AMD code is (R)-optimal precisely when the maximum success probability of the adversary over all $\delta \in G^*$ is equal to their average success probability.

Result

An AMD code is R-optimal $\Leftrightarrow$ e_δ is constant for all $\delta \in G^*$.

RWEDFs

We [HP] named the set systems corresponding to optimal AMD codes reciprocally-weighted external difference families (RWEDFs).

Definition

An $(n, m; k_1, \ldots, k_m; \ell)$-RWEDF is a collection of disjoint subsets $A_1, \ldots, A_m$ of a group G of order n, where $|A_i| = k_i$ for all $i \in \{1, \ldots, m\}$, with the property that:

$$\frac{1}{k_1} N_1(\delta) + \cdots + \frac{1}{k_m} N_m(\delta) = \ell$$

for all non-zero $\delta \in G$.

In the special case when sets A_i are of equal size, this becomes

$$N_1(\delta) + \cdots + N_m(\delta) = \text{ constant}$$

Have been studied (abelian): external difference families (EDFs).

EDF example

- Let $G = (\mathbb{Z}_{10}, +)$; take $A_1 = \{4, 7, 9\}$ and $A_2 = \{0, 2, 5\}$
- Differences from A_1 to A_2 are
  $\{4-0=4, 4-2=2, 4-5=-1=9,$
  $7-0=7, 7-2=5, 7-5=2,$
  $9-0=9, 9-2=7, 9-5=4\}$, ie $\{2, 2, 4, 4, 5, 7, 7, 9, 9\}$.
- Differences from A_2 to A_1 are their negatives, i.e. $\{1, 1, 3, 3, 5, 6, 6, 8, 8\}$.
- Union of all external differences=each nonzero element twice!
- For $\delta = 1$, the adversary's success probability is

$$\frac{1}{2}\left(\frac{N_1(\delta)}{|A_1|} + \frac{N_2(\delta)}{|A_2|}\right) = \frac{1}{2}\left(\frac{0}{3} + \frac{2}{3}\right) = \frac{1}{3}$$

- Same for any choice of $\delta \neq 0$.

Finite field EDF construction

Construction

Let G be the additive group of $GF(q)$, the finite field of order q, where q is a prime power congruent to 1 mod 4.
Let $A_1 = \{$the set of non-zero squares in $GF(q)\}$.
Let $A_2 = \{$the set of non-squares in $GF(q)\}$.
Then $\{A_1, A_2\}$ form a $(q, 2; \frac{q-1}{2}, \frac{q-1}{2}; 1)$-RWEDF (indeed an EDF).

Special case of cyclotomic method - using subgroups of the multiplicative group of a finite field to make EDFs in its additive group.

What about RWEDFs which are not EDFs?

Now we would like to construct examples of RWEDFs which have different set-sizes (i.e. are not EDFs).

Example

Let $G = \mathbb{Z}_{k_1 k_2 + 1}$. The sets

$$A_1 = \{0, 1, \ldots, k_1 - 1\} \text{ and } A_2 = \{k_1, 2k_1, \ldots, k_1 k_2\}$$

form a $(k_1 k_2 + 1, 2; k_1, k_2; \frac{1}{k_1} + \frac{1}{k_2})$-RWEDF.

Can prove: this gives an AMD code whose success probability is as small as possible when $m = 2$ for the given group size.

Difference sets

Definition

A difference set in a group G is a set $D \subseteq G$ such that, when we take all pairwise internal differences between the elements of D, every non-identity group element occurs a fixed number λ of times.

Result

Let G be a group of order n, and let $\mathcal{A} = \{A_1, A_2\}$ partition G. Then $\mathcal{A}$ is an RWEDF $\Leftrightarrow$ A_1 and A_2 are difference sets.

Example: Let $G = \mathbb{Z}_7$. Let $A_1 = \{1, 2, 4\}$ and $A_2 = \{0, 3, 5, 6\}$.
Then $\{A_1, A_2\}$ is a $(7, 2; 3, 4; \frac{7}{6})$-RWEDF.
For any δ, adversary's success probability is $\frac{7}{12}$.

Motivating questions for RWEDFs

Observe that all the examples we have seen so far have 2 sets.

Question

Can we get examples with more than 2 sets, ie $m > 2$?

The constant ℓ in the definition is in $\mathbb{Q}$ but not necessarily in $\mathbb{Z}$.

Question

Can we obtain constructions for RWEDFs with integer ℓ?

One way to guarantee integer ℓ would be if $k_i \mid N_i$ $(1 \leq i \leq m)$.

We must have $N_i(\delta) \leq k_i$ for $\delta \in G \setminus \{0\}$ - so this would mean $N_i(\delta) = 0$ or k_i.

Defining a New Property

Motivated by this condition for RWEDFs with integer ℓ, we define the following general property.

Let G be a finite group and let $\mathcal{A}$ be a collection $\{A_1, A_2, ..., A_m\}$ of disjoint subsets of G with sizes $k_1, k_2, \ldots k_m$ respectively.

Definition

We shall say $\mathcal{A}$ has the bimodal property if for all $\delta \in G^*$ we have $N_j(\delta) \in \{0, k_j\}$ for $j = 1, 2, \ldots, m$.

In other words: for each $\delta \in G^*$, either δ never occurs as a difference between A_i and some other A_j, or else for every $a_i \in A_i$ there is an $a_j \in A_j$ ($i \neq j$) s.t. $\delta = a_i - a_j$.

Not every collection of bimodal sets will be an RWEDF...

Example

Let $G = \mathbb{Z}_{10}$ and take $A_1 = \{1, 6\}$, $A_2 = \{3, 8\}$ and $A_3 = \{4, 9\}$. Then $A = \{A_1, A_2, A_3\}$ is bimodal but not an RWEDF. For $i = 3$ we have $N_3(1) = N_3(3) = N_3(6) = N_3(8) = 2 = k_3$ while

$$N_3(2) = N_3(4) = N_3(5) = N_3(7) = N_3(9) = 0.$$

Similar calculations for $N_1(\delta)$ and $N_2(\delta)$ confirm A has the bimodal property but 5 never occurs as an external difference.

...and not every RWEDF with integer ℓ will be bimodal - though some always will:

Result

An $(n, m; k_1, \ldots, k_m; \ell)$-RWEDF with $\ell \in \mathbb{Z}$ and $\{k_1, \ldots, k_m\}$ pairwise coprime is bimodal.

This opens up two quite distinct questions:

- Can families of sets with the bimodal property in finite (abelian) groups be algebraically characterized?
- Can we find bimodal families of sets which are RWEDFs?

Theoretical examples of bimodal sets

Result

Let H be a subgroup of an abelian group G. If $\mathcal{C} = \{C_1, \ldots, C_m\}$ is a collection of cosets of H, the $\mathcal{C}$ has the bimodal property.

Proof: For fixed i and $1 \leq j \leq m$ with $i \neq j$, the sets $C_i - C_j$ comprise $m-1$ distinct cosets of H. For any $\delta \in C_i - C_j$ and every $x \in C_i$, $\exists$ a unique $y \in C_j$ s.t. $x - y = \delta$. However, for any $\delta \in G \setminus \cup_{j \neq i}(C_i - C_j)$, δ occurs 0 times as a difference out of C_i.

Cosets are such a "natural" example that you may guess they are the only non-trivial collection of sets with the bimodal property, but in fact a much richer landscape emerges.

Internal difference group

Definition

Let A be a subset of a finite abelian group G. We define the internal difference group H of A to be the subgroup of G generated by all $x - y$ where $x, y \in A$, ie $H = \langle I(A) \rangle$.

The group H has the property that A is contained in a single coset of H, and is the smallest subgroup of G with this property.

Bimodal property and cosets

For disjoint subsets $\{A_1, \ldots, A_m\}$ of our group G, we will let:

- $A = \cup_{i=1}^{m} A_i$
- $B_i = A \setminus A_i$ for any $1 \leq i \leq m$
- $H_i = \langle I(A_i) \rangle$

Result

Let G be a finite abelian group and let $\mathcal{A} = \{A_1, \ldots, A_m\}$ be a collection of disjoint subsets of G. Then $\mathcal{A}$ has the bimodal property if and only if for each i the set B_i is a union of cosets of the subgroup H_i.

Subgroup partitions

Definition

If a finite group G has subgroups $S_1, \ldots, S_m$ with the property that $S_1 \setminus \{0\}, \ldots, S_m \setminus \{0\}$ partition $G \setminus \{0\}$, then the collection of subgroups is called a partition of G.

Example

Let $G = \mathbb{Z}_3 \times \mathbb{Z}_3$.
A partition of G is given by:
$S_1 = \{(0,0),(1,1),(2,2)\}, S_2 = \{(0,0),(0,1),(0,2)\},$
$S_3 = \{(0,0),(1,2),(2,1)\}, S_4 = \{(0,0),(1,0),(2,0)\}.$

Subgroup partitions give bimodal sets

Let $S_i{}^*$ denote $S_i \setminus \{0\}$.

Result

If a finite abelian group G has subgroups $S_1, \ldots, S_m$ forming a partition of G, then $\{S_1^*, \ldots, S_m^*\}$ has the bimodal property.

Proof: For each i, the internal difference group of S_i^* is S_i itself. So the union $\cup_{j \neq i} S_j^*$ is $G \setminus S_i$, a union of cosets of S_i.

Have seen: cosets and subgroup partitions - what is the general landscape for collections of bimodal sets?

Impressionistic idea of the situation

- Collections of bimodal sets in finite abelian groups are in some sense a "blend" of the coset and subgroup partition examples we've seen.
- Let $r_{\mathcal{A}}$ be the number of A_i with $|A_i| < |H_i|$.
  Key structural differences when $r_{\mathcal{A}} \geq 2$, $= 1$ and $= 0$.
- When $r_{\mathcal{A}} \geq 2$, the sets $A_1, \ldots, A_{r_{\mathcal{A}}}$ occur together in a very tightly-structured way: like an "inflated" group partition.
- This imposes considerable structure on the remaining members of $\mathcal{A}$ (coset part).
- The cases with $r_{\mathcal{A}} = 1$ and $= 0$ are comparable but have a simpler description.

Some technical points for our main structure result

- Recall $|A_i| \leq |H_i|$ for all $1 \leq i \leq m$.
  Wlog, we label the sets such that
  - $|A_i| < |H_i|$ for $i = 1, \dots, r_{\mathcal{A}}$
  - $|A_i| = |H_i|$ for $i = r_{\mathcal{A}} + 1, \dots, m$.
- Following literature, a collection of sets $F_1, \dots, F_k$ with the property that $F_i \cap F_j = D$ for all for all $i \neq j$ is said to be a *k-star with kernel D*.
- Helpful to shift to "canonical position": a translation guaranteeing that instead of cosets of certain subgroups, we are dealing with the subgroups themselves.

Structural result

Let $\mathcal{A} = \{A_1, \ldots, A_m\}$ be a bimodal collection of disjoint subsets of an abelian group G with $r_{\mathcal{A}} \geq 2$, in canonical position.

Result

- *The internal difference groups $H_1, \ldots, H_{r_{\mathcal{A}}}$ form an $r_{\mathcal{A}}$-star with kernel $D_{\mathcal{A}}$ (a subgroup of G), and for each i with $1 \leq i \leq r_{\mathcal{A}}$ we have $A_i = H_i \setminus D_{\mathcal{A}}$.*
- *Any set A_i with $i > r_{\mathcal{A}}$ is a coset of a subgroup of $D_{\mathcal{A}}$*
- *If H denotes the group $H_1 + H_2 + \cdots + H_{r_{\mathcal{A}}}$, then $H \setminus D_{\mathcal{A}}$ is contained in A . Furthermore, the sets in $\mathcal{A}$ can be labelled such that for some k with $r_{\mathcal{A}} \leq k \leq m$ we have that $H \setminus D_{\mathcal{A}}$ is partitioned by $A_1, \ldots, A_k$.*
- *If $k < m$ then the sets A_i with $i > k$ arise from a subdivision of cosets of H.*

We also have the other way round...

Let G be an abelian group, and for $t \geq 2$ let $H_1, \ldots, H_t$ be distinct subgroups of G forming a t-star with kernel D, such that $|H_i : D| > 2$ for i with $1 \leq i \leq t$.
Let $H = H_1 + \cdots + H_t$.

Result

Let $\mathcal{A}$ consist of the following subsets of G:

- *all subsets of the form $A_i = H_i \setminus D$ for i with $1 \leq i \leq t$;*
- *all cosets of D that are subsets of H, but are not in $\cup_{i=1}^{t} H_i$;*
- *for any number of cosets of H, all the cosets of D that lie within those cosets of H.*

Then $\mathcal{A}$ is a bimodal collection of subsets of G with $r_{\mathcal{A}} = t$ in canonical position.

Return to RWEDFs...

Returning to RWEDFs, we can prove the following for any (not necessarily abelian) finite group:

Result

If a finite group G of order n has subgroups $S_1, \ldots S_m$ forming a partition of G, then $\{S_1^*, \ldots, S_m^*\}$ is a (bimodal) RWEDF.

- This gives a wealth of new RWEDF/EDF examples, in both abelian and non-abelian groups.

Groups possessing subgroup partitions

From the literature, groups which admit a subgroup partition include:

- elementary abelian p-groups of order $\geq p^2$, for p prime
- Frobenius groups (eg dihedral group D_{2n} with n odd)
- groups of Hughes-Thompson type
- groups isomorphic to $PGL(2, p^h)$ with p an odd prime

RWEDF example from subgroup partition

Example

- *Let* $G = \mathbb{Z}_3 \times \mathbb{Z}_3$.
- *We use the subgroup partition from the earlier slide, removing the zero element.*
- *Let* $A_1 = \{(1,1),(2,2)\}$, $A_2 = \{(0,1),(0,2)\}$, $A_3 = \{(1,2),(2,1)\}$ *and* $A_4 = \{(1,0),(2,0)\}$.
- *For non-zero* $\delta \in G$, $N_i(\delta) = 2$ *for* $\delta \notin A_i$ *and* $N_i(\delta) = 0$ *for* $\delta \in A_i$ *(for each* $1 \leq i \leq 4$*).*
- $\mathcal{A}$ *forms a* $(9,4;2,2,2,2;3)$*-RWEDF (indeed, this is an EDF).*

This is an example of a more general construction we have in elementary abelian p-groups using vector space partitions.

Nonabelian RWEDF partition example

Example

Let n be odd, and let D_{2n} be the dihedral group given by the presentation

$$\langle r, s | r^n = 1, s^2 = 1, rs = sr^{-1} \rangle$$

A partition is given by $S_i = \langle sr^{i-1} \rangle$ for $1 \leq i \leq n$ and $S_{n+1} = \langle r \rangle$.
Here $|S_1| = \cdots = |S_n| = 2$ and $|S_{n+1}| = n$.
For D_{10} this yields a $(10, 6; 1, 1, 1, 1, 1, 4; 5)$-RWEDF.

Nonabelian RWEDF partition example

Example

Let G be the Heisenberg group *modulo 3, ie the group of 3×3 upper triangle matrices with entries from $GF(3)$ that have 1s on the main diagonal.*

Each element of G has the form $\begin{pmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix}$ *and each non-identity element has order 3.*

$|G| = 27$ *and its order 3 subgroups partition its non-identity elements.*

This will give an EDF with 13 sets of size 2.

A stronger security model

Question: is it always realistic to assume that an adversary will not know which message (source) is being sent?

We may wish to consider a stronger security model, in which the adversary knows the source before they choose their δ.

Stronger security model

Strong AMD code

Encoder: chooses a source $s \in S$

Adversary: is given source s

Adversary: chooses some $\delta \in G \setminus \{0\}$

Encoder: source is encoded by E to $g \in A(s)$

Adversary: g is replaced by $g' = g + \delta$

Adversary wins if $g' \in A(s')$ for some $s' \neq s$.

In a strong AMD code, the adversary learns s before choosing δ.

Strong AMD codes and set systems

What mathematical structures correspond to optimal strong AMD codes?

At present: strong EDFs are used.

These require a condition for each possible i:

Definition

A strong external difference family in an abelian group G of order n is a collection of disjoint sets $A_1, \ldots, A_m$ of G, each of size k, such that when we take all external differences from any A_i to $\cup_{j \neq i} A_j$, every non-identity group element occurs a fixed number ℓ of times.

We write this as an (n, m, k, ℓ)-SEDF.

Examples of SEDFs

In fact, we have seen examples of SEDFs earlier in the talk:

- In $G = \mathbb{Z}_5$, the sets $\{1, 4\}$ and $\{2, 3\}$ form an SEDF.
- Let G be the additive group of $GF(q)$ where q is a prime power congruent to 1 mod 4; the set of non-zero squares and the set of non-squares form an SEDF.
- In $G = \mathbb{Z}_{k^2+1}$, the sets

  $$A_1 = \{0, 1, \ldots, k-1\} \text{ and } A_2 = \{k, 2k, \ldots, k^2\}$$

  form a $(k^2+1, 2, k, 1)$-SEDF.

The SEDF landscape: existence

SEDFs are known to exist for the following (n, m, k, ℓ):

(a) $(k^2+1, 2, k, 1)$: $G = \mathbb{Z}_{k^2+1}$, Paterson/Stinson

(b) $(v, 2, \frac{v-1}{2}, \frac{v-1}{4})$ where $v \equiv 1 \bmod 4$ and an appropriate partial difference set exists: Davis/Huczynska/Mullen and Huczynska/Paterson

(c) $(q, 2, \frac{q-1}{4}, \frac{q-1}{16})$ where $q = 16t^2 + 1$ is a prime power and $G = GF(q)$: Bao/Wei/Zhang

(d) $(q, 2, \frac{q-1}{6}, \frac{q-1}{36})$ where $q = 108t^2 + 1$ is a prime power and $G = GF(q)$: Bao/Wei/Zhang

First SEDF with $m > 2$

Until the start of 2018, no SEDFs were known with $m \neq 2$.
Then the first with $m > 2$ was found - independently by two sets of authors.

$(243, 11, 22, 20)$-SEDF in $\mathbb{Z}_3^5$

- Cyclotomic construction [Wen,Yang,Feng]
- Action of M_{11} on $\mathrm{PG}(4, 3)$ [Jedwab,Li]

This is still the only known SEDF with more than 2 sets!

Non-abelian SEDFs

One theme of the bimodal work was the emergence of non-abelian RWEDF examples.

Recently, we [HJN] obtained the first construction for a family of non-abelian SEDFs:

Theorem

Let $k > 1$ be odd. In D_{k^2+1}, the dihedral group of order $n = k^2 + 1$, there exists a $(k^2+1, 2, k, 1)$-SEDF in G. Specifically, in

$$\langle r, s | r^{n/2} = 1, s^2 = 1, rs = sr^{-1} \rangle$$

we can take $\{A_1, A_2\}$ where

- $A_1 = \{r^i : 0 \leq i \leq \frac{k-1}{2}\} \cup \{sr^j : 0 \leq j \leq \frac{k-3}{2}\}$.
- $A_2 = \{r^{ik} : 1 \leq i \leq \frac{k-1}{2}\} \cup \{sr^{\frac{k(2j+1)-1}{2}} : 0 \leq j \leq \frac{k-1}{2}\}$.

Open questions

There are many avenues to explore further in this area.

- Beyond group partitions, which collections of bimodal sets guarantee RWEDFs?
- Obtain a combinatorial characterization of RWEDFs with integer ℓ.
- Constructions for RWEDFs with integer ℓ which are not bimodal?
- Fine-tune our constructions to yield smallest possible success probabilities.
- The strong model for RWEDFs.
- Further constructions in nonabelian groups.
- Specific connections with Frobenius groups?

Thank you for listening!