Classifications of (some) Partial MDS Codes

Anna-Lena Horlemann-Trautmann

University of St. Gallen, Switzerland

July 11th, 2019
SIAM Conference on Applied Algebraic Geometry
Bern, Switzerland

Joint work with Alessandro Neri, Tristram Bogart, David Karpuk, Mauricio Velasco.

1 Distributed Storage and PMDS Codes

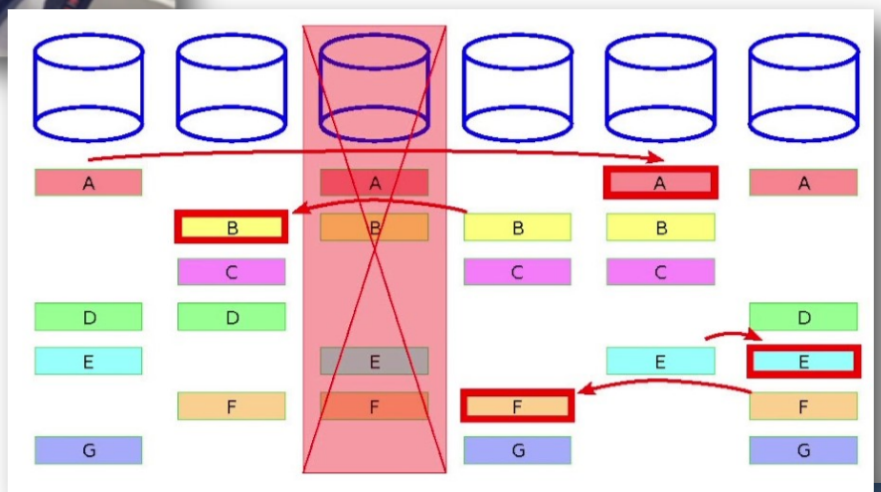
2 Classification Results for $s = 1$

3 (Classification) Results for $\ell = 2$

4 Conclusions



Distributed Storage





Graphics from `https://people.eecs.berkeley.edu/~kannanr/project_dss.html`,
`https://www.youtube.com/watch?v=8XBNP6U8IZs`,
`https://www.slashgear.com/the-best-cloud-storage-of-2017-an-overview-of-providers-14488448/`

Distributed Storage

- Distributed storage systems: store file/codeword over several servers/disks
- Traditionally: replication code (very high storage overhead)
- Theoretical alternative: erasure codes (very high repair cost)
- Locally repairable codes: find a balance between the two

Locality

- Locality of node/coordinate: # nodes needed for recovering
- Locality of code: maximum of all node localities

Distributed Storage

- Distributed storage systems: store file/codeword over several servers/disks
- Traditionally: replication code (very high storage overhead)
- Theoretical alternative: erasure codes (very high repair cost)
- Locally repairable codes: find a balance between the two

Locality

- Locality of node/coordinate: # nodes needed for recovering
- Locality of code: maximum of all node localities

Our setting: disjoint (or symmetric) locality groups

Definition (Partial MDS Codes)

Let $B_i \in \mathbb{F}_q^{k\times(r_i+\ell)}$ and

$$G = (B_1 \mid \cdots \mid B_m) \in \mathbb{F}_q^{k\times n}.$$

Then G generates a $[n, k, \ell; r_1, \ldots, r_m]$-*partial-MDS (PMDS) code* (with locality ℓ) if

- for $i \in \{1, \ldots, m\}$ the row space of B_i is a $[r_i + \ell, \ell]$-MDS code, and
- any collection of ℓ columns per block generate a $[m\ell, k]$-MDS code.

(Blaum-Hafner-Hetzler '13, HT-Neri '19)

Lemma

A $[n, k, \ell; r_1, \ldots, r_m]$-PMDS code can correct any r_i erasures in the i-th block (simultaneously) plus $s := m\ell - k$ additional erasures anywhere in the code.

Lemma

A $[n, k, \ell; r_1, \ldots, r_m]$-PMDS code can correct any r_i erasures in the i-th block (simultaneously) plus $s := m\ell - k$ additional erasures anywhere in the code.

We know:

- PMDS codes are exactly maximally recoverable (MR) codes, if the locality groups are disjoint.
- PMDS codes exist if $q > \binom{n-1}{k-1}$.
- Constructions based on Reed-Solomon and Gabidulin codes with $q \geq n^{m\ell}$.
- Construction for $s = 1$ and $r_1 = \cdots = r_m$ with $q \geq \ell + r$.

1 Distributed Storage and PMDS Codes

2 Classification Results for $s = 1$

3 (Classification) Results for $\ell = 2$

4 Conclusions

Theorem

The matrix $H \in \mathbb{F}_q^{(n-k)\times n}$ *is a parity check matrix of a* $[n, k, \ell; r_1, \ldots, r_m]$*-PMDS code of length* $n = m\ell + \sum_i r_i$*, dimension* $k = m\ell - 1$ *and locality* ℓ *over* $\mathbb{F}_q$*:*

$$H = \left(\begin{array}{cccc} H_1 & 0 & \ldots & 0 \\ 0 & H_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & H_m \\ \hline v_1 & v_2 & \ldots & v_m \end{array}\right),$$

where $H_i \in \mathbb{F}_q^{r_i\times(\ell+r_i)}$ *is a parity check matrix of a* $[\ell + r_i, \ell]$*-MDS code, and* $v_i \in \mathbb{F}_q^{1\times(\ell+r_i)}$ *is such that* $\begin{pmatrix} H_i \\ v_i \end{pmatrix}$ *is a parity check matrix of a* $[\ell + r_i, \ell - 1]$*-MDS code.*

Example

The parity check matrix

$$H = \left(\begin{array}{ccc|ccc} 2 & 1 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 2 & 2 & 1 \\ \hline 1 & 1 & 0 & 1 & 2 & 0 \end{array}\right).$$

and the corresponding generator matrix

$$G = \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 2 & 0 & 1 & 1 \\ \hline 0 & 0 & 0 & 1 & 1 & 2 \end{array}\right)$$

define a PMDS code with $[n, k, \ell; r_1, r_2] = [6, 3, 2; 1, 1]$ over $\mathbb{F}_3$.

Lemma

A $[n, k]$-MDS code containing a $[n, k-1]$-MDS code exists if and only if there exists a $[n+1, k]$-MDS code.

Lemma

A $[n, k]$-MDS code containing a $[n, k-1]$-MDS code exists if and only if there exists a $[n+1, k]$-MDS code.

Theorem (HT-Neri '19)

1. *For any integers $m \geq 2$ and $\ell, r_1, \ldots, r_m \geq 1$ there exists a $[n, k = m\ell - 1, \ell; r_1, \ldots, r_m]$-PMDS code over any field $\mathbb{F}_q$ with $q \geq \max_i\{r_i\} + \ell$.*

Lemma

A $[n, k]$-MDS code containing a $[n, k-1]$-MDS code exists if and only if there exists a $[n+1, k]$-MDS code.

Theorem (HT-Neri '19)

1. *For any integers $m \geq 2$ and $\ell, r_1, \ldots, r_m \geq 1$ there exists a $[n, k = m\ell - 1, \ell; r_1, \ldots, r_m]$-PMDS code over any field $\mathbb{F}_q$ with $q \geq \max_i\{r_i\} + \ell$.*
2. *If there exists $h \in \mathbb{N}$ such that $\max_i\{r_i\} + \ell = 2^h + 1$ and $\ell \in \{3, 2^h - 1\}$, then there exists an $[n, k = m\ell - 1, \ell; r_1, \ldots, r_m]$-PMDS code over $\mathbb{F}_q$ with $q = \max_i\{r_i\} + \ell - 1$.*

Lemma

A $[n, k]$-MDS code containing a $[n, k-1]$-MDS code exists if and only if there exists a $[n+1, k]$-MDS code.

Theorem (HT-Neri '19)

1. *For any integers $m \geq 2$ and $\ell, r_1, \ldots, r_m \geq 1$ there exists a $[n, k = m\ell - 1, \ell; r_1, \ldots, r_m]$-PMDS code over any field $\mathbb{F}_q$ with $q \geq \max_i\{r_i\} + \ell$.*
2. *If there exists $h \in \mathbb{N}$ such that $\max_i\{r_i\} + \ell = 2^h + 1$ and $\ell \in \{3, 2^h - 1\}$, then there exists an $[n, k = m\ell - 1, \ell; r_1, \ldots, r_m]$-PMDS code over $\mathbb{F}_q$ with $q = \max_i\{r_i\} + \ell - 1$.*
3. *If $\ell = 1$, then there exists a $[n, k = m - 1, 1; r_1, \ldots, r_m]$-PMDS code over $\mathbb{F}_q$ with $q \geq 2$.*

- The previous results are quite straight-forward.
- However, to show that all PMDS code with $s = 1$ are of this form, is more involved.
- To show this we change to the generator matrix representation (since puncturing can be represented there).

Theorem (HT-Neri '19)

A linear code over $\mathbb{F}_q$ of length $n = m\ell + \sum_{i=1}^{m} r_i$ and dimension $k = m\ell - 1$ is a $[n, k, \ell; r_1, \ldots, r_m]$-PMDS code if and only if it has a generator matrix of the form

$$G = \left(\begin{array}{cccc|c} B_1 & 0 & \ldots & 0 & M_1 \\ 0 & B_2 & \ldots & 0 & M_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & B_{m-1} & M_{m-1} \\ \hline 0 & 0 & \ldots & 0 & A \end{array}\right)$$

where the $(B_i \quad | M_i[col_\ell])$ are generator matrices of a $[\ell + r_i + 1, \ell]$-MDS code and

$$\left(\begin{array}{c|c} 1 & M_i[row_1] \\ \hline 0 & A \end{array}\right)$$

generates a $[\ell + r_m + 1, \ell]$-MDS code.

Theorem (HT-Neri '19)

A $[n, k = m\ell - 1, \ell; r_1, \ldots, r_m]$-PMDS code over $\mathbb{F}_q$ exists if and only if, for $i = 1, \ldots, m$, there exists a $[\ell + \max_i\{r_i\} + 1, \ell]$-MDS code over $\mathbb{F}_q$.

Theorem (HT-Neri '19)

A $[n, k = m\ell - 1, \ell; r_1, \ldots, r_m]$-PMDS code over $\mathbb{F}_q$ exists if and only if, for $i = 1, \ldots, m$, there exists a $[\ell + \max_i\{r_i\} + 1, \ell]$-MDS code over $\mathbb{F}_q$.

Corollary (HT-Neri '19)

Assuming that the MDS-conjecture is correct, we have:

1. *If there exists a $[n, k = m\ell - 1, \ell; r_1, \ldots, r_m]$-PMDS code over $\mathbb{F}_q$ such that the parameters are not included in Case 2 and $\ell > 1$, then $q \geq \max_i\{r_i\} + \ell$.*
2. *If there exists a $[n, k = m\ell - 1, \ell; r_1, \ldots, r_m]$-PMDS code over $\mathbb{F}_q$ such that $\ell \in \{3, 2^h - 1\}$ and $\max_i\{r_i\} + \ell = 2^h + 1$ (for some $h > 1$), then $q \geq 2^h = \max_i\{r_i\} + \ell - 1$.*

1 Distributed Storage and PMDS Codes

2 Classification Results for $s = 1$

3 (Classification) Results for $\ell = 2$

4 Conclusions

What does locality $\ell = 2$ mean?

- In each block we want to recover the whole (block) code from any two columns.

What does locality $\ell = 2$ mean?

- In each block we want to recover the whole (block) code from any two columns.
- Idea: Let columns of block be projective line – then we recover the whole line from any two points.
- Then we still want to correct the s global erasures – for this we need to take some of the points out of the code (generator matrix).

What does locality $\ell = 2$ mean?

- In each block we want to recover the whole (block) code from any two columns.
- Idea: Let columns of block be projective line – then we recover the whole line from any two points.
- Then we still want to correct the s global erasures – for this we need to take some of the points out of the code (generator matrix).

$\Longrightarrow$ In fact, every PMDS with $\ell = 2$ will be like this!

More rigorously...

Theorem

Let $B_i \in \mathbb{F}_q^{2\times(r_i+\ell)}$ *and* $A \in \mathbb{F}_q^{I(s \text{ odd})\times\sum_{i=0}^{\lfloor \frac{s}{2} \rfloor}(r_{m-i}+\ell)}$. *Any* $[n, k, 2; r_1, \ldots, r_m]$*-PMDS code has a generator matrix of the form*

$$G = \left(\begin{array}{cccc|c} B_1 & 0 & \ldots & 0 & M_1 \\ 0 & B_2 & \ldots & 0 & M_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & B_{m-\lfloor \frac{s}{2} \rfloor} & M_{m-\lfloor \frac{s}{2} \rfloor} \\ \hline 0 & 0 & \ldots & 0 & A \end{array}\right)$$

such that the row space per block is a $[r_i + 2, 2]$*-MDS code, and the collection of any two columns per block generates an* $[2m, 2m - 2s]$*-MDS code.*

Lemma

Any $[n, 2]$-MDS code is the image of the set of linear maps, evaluated in points on a projective line in $\mathbb{P}^{n'-1}$, with $n' \geq n$.

Lemma

Any $[n, 2]$-MDS code is the image of the set of linear maps, evaluated in points on a projective line in $\mathbb{P}^{n'-1}$, with $n' \geq n$.

Theorem (Bogart-HT-Karpuk-Neri-Velasco '19-20)

The columns of the generator matrix of a $[n, k, 2; r_1, \ldots, r_m]$-PMDS code are points in $\mathbb{P}^{k-1}$, such that the columns of each block lie on a projective line, and

1. *for $s = 1$, any $m - 1$ of the lines, plus any point from the remaining line, generate the whole $\mathbb{P}^{k-1}$;*
2. *for $s = 2$*
   - *any $m - 1$ of the lines generate the whole $\mathbb{P}^{k-1}$, and*
   - *any $m - 2$ lines, plus any point each from the remaining two lines, generate the whole $\mathbb{P}^{k-1}$.*

Back to our example with $s = 1$

The generator matrix

$$G = \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 2 & 0 & 1 & 1 \\ \hline 0 & 0 & 0 & 1 & 1 & 2 \end{array}\right)$$

defines a PMDS code with $[n, k, \ell; r_1, r_2] = [6, 3, 2; 1, 1]$ over $\mathbb{F}_3$.

- For both blocks, the three columns are points on a projective line in $\mathbb{P}^2(\mathbb{F}_3)$.
- Any point of the left block is not on the line of the second block, and vice versa.

Back to our example with $s = 1$

The generator matrix

$$G = \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 2 & 0 & 1 & 1 \\ \hline 0 & 0 & 0 & 1 & 1 & 2 \end{array}\right)$$

defines a PMDS code with $[n, k, \ell; r_1, r_2] = [6, 3, 2; 1, 1]$ over $\mathbb{F}_3$.

- For both blocks, the three columns are points on a projective line in $\mathbb{P}^2(\mathbb{F}_3)$.
- Any point of the left block is not on the line of the second block, and vice versa.
- We cannot add the remaining points on the lines, otherwise we loose the global MDS property.

Example with $s = 2$

The generator matrix

$$G = \left(\begin{array}{ccc|ccc|ccc} 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 & 5 & 6 & 7 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 2 & 3 & 4 & 5 & 6 & 7 \end{array}\right)$$

defines a PMDS code with $[n, k, \ell; r_1, r_2, r_3] = [9, 4, 2; 1, 1, 1]$ over $\mathbb{F}_{11}$.

Idea of general construction for $s = \ell = 2$

- Choose m lines in general position in $\mathbb{P}^{k-1}$.
- Construct m-sets, containing a point from each line, that are collinear/coplanar. These sets partition the set of all considered points.
- Keep one point per set, distributed evenly (or as desired) across the lines.

The resulting matrix generates a $[q+1, 2(m-1), 2; r_1, \ldots, r_m]$-PMDS code.

Theorem (Bogart-HT-Karpuk-Neri-Velasco '19-20)

$[n, 2(m-1), 2; r_1, \ldots, r_m]$-PMDS codes exist over $\mathbb{F}_q$, if $q \geq n - 1$.

Theorem (Bogart-HT-Karpuk-Neri-Velasco '19-20)

$[n, 2(m-1), 2; r_1, \ldots, r_m]$-PMDS codes exist over $\mathbb{F}_q$, if $q \geq n - 1$.

Strong conjecture: Field size cannot be smaller.

Theorem (Bogart-HT-Karpuk-Neri-Velasco '19-20)

$[n, 2(m-1), 2; r_1, \ldots, r_m]$-PMDS codes exist over $\mathbb{F}_q$, if $q \geq n - 1$.

Strong conjecture: Field size cannot be smaller.

Weak conjecture: All PMDS codes with $\ell = s = 2$ are of this type (classification).

1 Distributed Storage and PMDS Codes

2 Classification Results for $s = 1$

3 (Classification) Results for $\ell = 2$

4 Conclusions

Conclusion

- PMDS codes are known to exist over large fields.
- For $s = 1$ (one global parity) we gave a complete classification, showing that they exist if and only if the field size is at least the length of the largest block.
- For locality $\ell = 2$ we show that the columns of the generator matrix, seen as points in $\mathbb{P}^{k-1}$, lie on projective lines per block.
- We use this to derive a general construction, over fields of size $q \geq n - 1$.
- Current work: generalization to larger s.
- Note: $s = 0$ and $\ell = 1$ give trivial results.

Conclusion

- PMDS codes are known to exist over large fields.
- For $s = 1$ (one global parity) we gave a complete classification, showing that they exist if and only if the field size is at least the length of the largest block.
- For locality $\ell = 2$ we show that the columns of the generator matrix, seen as points in $\mathbb{P}^{k-1}$, lie on projective lines per block.
- We use this to derive a general construction, over fields of size $q \geq n - 1$.
- Current work: generalization to larger s.
- Note: $s = 0$ and $\ell = 1$ give trivial results.

Thank you for your attention!
Questions? – Comments?

