Codes with Locality in the Rank and Subspace Metrics

Swanand Kadhe
UC Berkeley

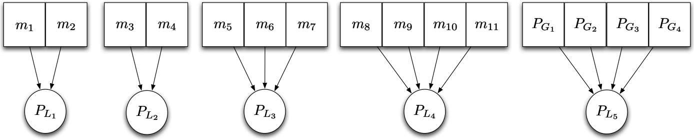
Joint work with

Salim El Rouayheb (Rutgers)
Iwan Duursma (UIUC)
Alex Sprintson (Texas A&M)

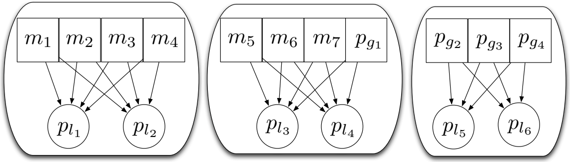
ITA '18
Feb 16, 2018

Locality of a Code

- Consider an (n, k, d) code $\mathcal{C}$ over $\mathbb{F}_q$
- Locality r: any codeword symbol can be recovered from some other r symbols of $\mathcal{C}$



Local codes have minimum Hamming distance of 2



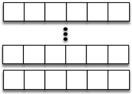
Local codes have minimum Hamming distance of 3

Gopalan *et al.* '12, Papailiopoulos-Dimakis '14, Prakash *et al.* '14, Tamo-Barg '14, Huang *et al.* '16, Gopalan *et al.* '17, ..., ..., ...

Choosing a Metric

Conventional Codes

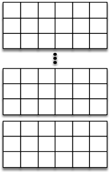
Codewords: vectors



Hamming distance

Rank-metric Codes

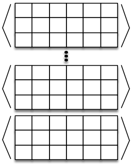
Codewords: matrices



Rank distance

Subspace Codes

Codewords: subspaces

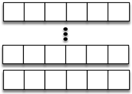


Subspace distance

Choosing a Metric

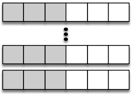
Conventional Codes

Codewords: vectors



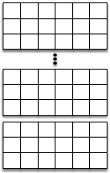
Hamming distance

Locality:



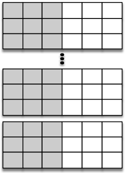
Rank-metric Codes

Codewords: matrices



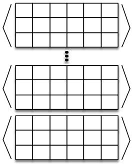
Rank distance

Locality:



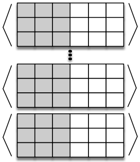
Subspace Codes

Codewords: subspaces



Subspace distance

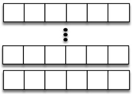
Locality:



Choosing a Metric

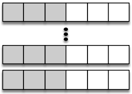
Conventional Codes

Codewords: vectors



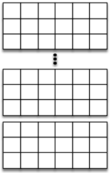
Hamming distance

Locality:



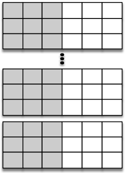
Rank-metric Codes

Codewords: matrices



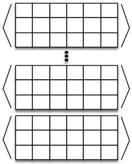
Rank distance

Locality:



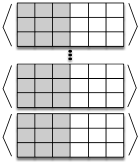
Subspace Codes

Codewords: subspaces



Subspace distance

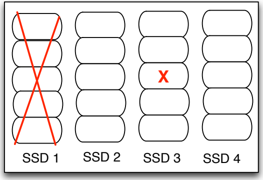
Locality:



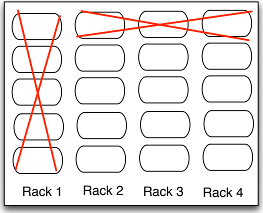
We focus on locality in rank and subspace metrics

Why to Consider Locality in Rank and Subspace Metrics?

- Mixed and correlated failures
  - Mixed failures: entire drive (node) plus a few blocks fail
  - Correlated failures: a bunch of nodes fail simultaneously



SSD Array



Data center

Example: Mixed failure in a solid state drive (SSD) array, and a correlated failure in a data center

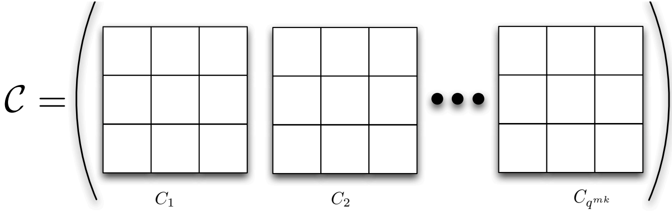
- Distributed storage over a network introducing errors and erasures
  - Repairing a failed node from a subset of nodes
  - Downloading partial data by connecting to only a small subset of nodes

Our Contributions

1. Notions of rank-locality and subspace-locality
2. A Singleton-like upper bound on the minimum rank-distance for codes with rank-locality
3. Construct a class of distance-optimal codes with rank-locality building up on Tamo-Barg construction
4. Obtain a class of codes with subspace-locality by lifting rank metric codes

Rank-Metric Codes

- A rank-metric code $\mathcal{C}$ is a non-empty subset of $\mathbb{F}_q^{m\times n}$ of size q^{mk} endowed with rank-distance metric
  $d_R(A, B) = \text{rank}(A - B)$ [Delsarte '78, Gabidulin '85, Roth '91]



- Maximum rank-distance (MRD) codes are analogues of the maximum distance separable (MDS) codes in the Hamming metric
  - MRD codes achieve the Singleton bound for the rank-metric codes

$$|\mathcal{C}| \leqslant q^{\max\{n,m\}(\min\{n,m\}-d+1)}$$

Gabidulin Codes

Rank-metric analogues of Reed-Solomon codes

- Let $P = \{p_1, \cdots, p_n\}$ be a set of n elements in $\mathbb{F}_{q^m}$ that are linearly independent over $\mathbb{F}_q$ $(m \geqslant n)$
- Let $G_{\mathbf{m}}(x) \in \mathbb{F}_{q^m}[x]$ denote the linearized polynomial of q-degree at most $k-1$ with coefficients $\mathbf{m}$ as follows.

$$G_{\mathbf{m}}(x) = \sum_{j=0}^{k-1} m_j x^{q^j}, \qquad G = \begin{bmatrix} p_1 & p_2 & \cdots & p_n \\ p_1^q & p_2^q & \cdots & p_n^q \\ p_1^{q^2} & p_2^{q^2} & \cdots & p_n^{q^2} \\ \vdots & \vdots & \ddots & \vdots \\ p_1^{q^{k-1}} & p_2^{q^{k-1}} & \cdots & p_n^{q^{k-1}} \end{bmatrix}$$

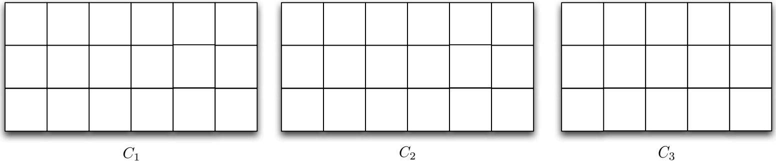
- Gabidulin code is obtained by the following evaluation map

$$\begin{aligned} \mathrm{Enc} : \mathbb{F}_{q^m}^k &\to \mathbb{F}_{q^m}^n \\ \mathbf{m} &\mapsto \{G_{\mathbf{m}}(p_i), p_i \in P\} \end{aligned}$$

(r, δ) Rank-Locality

- An $(m \times n, k)$ rank-metric code $\mathcal{C}$ is said to have (r, δ) rank-locality if for each column $i \in [n]$ of the codeword matrix, there exists a set of columns $\Gamma(i) \subset [n]$ such that
  1. $i \in \Gamma(i)$,
  2. $|\Gamma(i)| \leqslant r + \delta - 1$, and
  3. $d_R\left(\mathcal{C}|_{\Gamma(i)}\right) \geqslant \delta$,

  where $\mathcal{C}|_{\Gamma(i)}$ is the restriction of $\mathcal{C}$ on the columns indexed by $\Gamma(i)$
- The code $\mathcal{C}|_{\Gamma(i)}$ is said to be the local code associated with the i-th column



Rank-metric code with $(4, 3)$ rank-locality: local codes C_1, C_2, and C_3 are rank-metric codes with rank-distance at least 3

Rank-Locality: Minimum Distance Bound

Theorem: For a rank-metric code $\mathcal{C} \subseteq \mathbb{F}_q^{m\times n}$ of cardinality q^{mk} with (r,δ) rank-locality, it holds that

$$d_R(\mathcal{C}) \leqslant n - k + 1 - \left(\left\lceil \frac{k}{r} \right\rceil - 1\right)(\delta - 1).$$

Rank-Locality: Minimum Distance Bound

Theorem: For a rank-metric code $\mathcal{C} \subseteq \mathbb{F}_q^{m\times n}$ of cardinality q^{mk} with (r, δ) rank-locality, it holds that

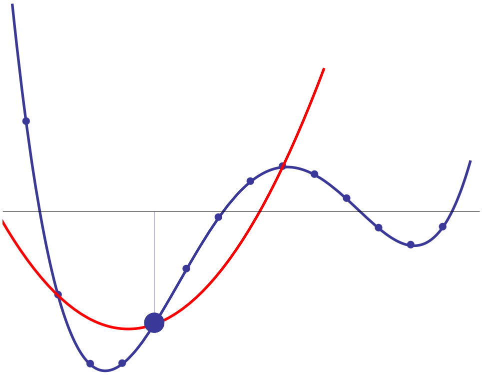
$$d_R(\mathcal{C}) \leqslant n - k + 1 - \left(\left\lceil \frac{k}{r} \right\rceil - 1\right)(\delta - 1).$$

Proof Sketch:

- Proof follows from the Singleton-like bound for the Hamming metric by [Prakash *et al.* '13, Rawat *et al.* '14]

Rank-Locality: Code Construction

We build upon the construction of [Tamo-Barg '14]



- Intuition: What if we can interpolate low degree polynomials to recover an erased symbol?
- For the rank-locality, we need to use linearized polynomials

Rank-Locality: Code Construction

Assume: $r \mid k$, $(r+\delta-1) \mid n$, $n \mid m$, $\mu := n/(r+\delta-1)$, $q \geqslant 2$

- Encoding Linearized Polynomial:
  - Given k information symbols m_{ij}, $i = 0, \ldots, r-1; j = 0, \ldots, \frac{k}{r}-1$, define the encoding polynomial as

$$G_{\mathbf{m}}(x) = \sum_{i=0}^{r-1} \sum_{j=0}^{\frac{k}{r}-1} m_{ij} x^{q^{(r+\delta-1)j+i}}.$$

Rank-Locality: Code Construction

Assume: $r \mid k$, $(r+\delta-1) \mid n$, $n \mid m$, $\mu := n/(r+\delta-1)$, $q \geqslant 2$

- Encoding Linearized Polynomial:
  - Given k information symbols m_{ij}, $i = 0, \ldots, r-1; j = 0, \ldots, \frac{k}{r} - 1$, define the encoding polynomial as

$$G_{\mathbf{m}}(x) = \sum_{i=0}^{r-1} \sum_{j=0}^{\frac{k}{r}-1} m_{ij} x^{q^{(r+\delta-1)j+i}}.$$

- Evaluation Points:
  - $\{\alpha_1, \ldots, \alpha_{r+\delta-1}\}$: basis of $\mathbb{F}_{q^{r+\delta-1}}$ as a vector space over $\mathbb{F}_q$
  - $\{\beta_1, \ldots, \beta_\mu\}$: basis of $\mathbb{F}_{q^n}$ as a vector space over $\mathbb{F}_{q^{r+\delta-1}}$
  - Evaluation points are $P_1, P_2, \cdots, P_\mu$, where $P_j = \{\alpha_i \beta_j, 1 \leqslant i \leqslant r+\delta-1\}$

Rank-Locality: Code Construction

Assume: $r \mid k$, $(r+\delta-1) \mid n$, $n \mid m$, $\mu := n/(r+\delta-1)$, $q \geqslant 2$

- Encoding Linearized Polynomial:
  - Given k information symbols m_{ij}, $i = 0, \ldots, r-1; j = 0, \ldots, \frac{k}{r}-1$, define the encoding polynomial as

$$G_{\mathbf{m}}(x) = \sum_{i=0}^{r-1} \sum_{j=0}^{\frac{k}{r}-1} m_{ij} x^{q^{(r+\delta-1)j+i}}.$$

- Evaluation Points:
  - $\{\alpha_1, \ldots, \alpha_{r+\delta-1}\}$: basis of $\mathbb{F}_{q^{r+\delta-1}}$ as a vector space over $\mathbb{F}_q$
  - $\{\beta_1, \ldots, \beta_\mu\}$: basis of $\mathbb{F}_{q^n}$ as a vector space over $\mathbb{F}_{q^{r+\delta-1}}$
  - Evaluation points P and their partition $(P_1, P_2, \cdots, P_\mu)$ is given as $P_j = \{\alpha_i \beta_j, 1 \leqslant i \leqslant r+\delta-1\}$
- Codeword is the evaluations of $G_{\mathbf{m}}(x)$ on points in P, *i.e.*, $\mathbf{c} = (G_{\mathbf{m}}(\gamma), \gamma \in P)$

Proposed Construction: Example

$n = 9, k = 4, r = 2, \delta = 2$. Set $q = 2$ and $m = n$

ω: primitive element of $\mathbb{F}_{2^9}$

- Define the encoding polynomial as

$$G_{\mathbf{m}}(x) = m_{00}x^{2^0} + m_{01}x^{2^3} + m_{10}x^{2^1} + m_{11}x^{2^4}.$$

- Obtain the evaluation points as
  - $\{1, \omega^{73}, \omega^{146}\}$: a basis of $\mathbb{F}_{2^3}$ over $\mathbb{F}_2$
  - $\{1, \omega^{309}, \omega^{107}\}$: a basis of $\mathbb{F}_{2^9}$ over $\mathbb{F}_{2^3}$

$$P = \{\{1, \omega^{73}, \omega^{146}\}, \{\omega^{309}, \omega^{382}, \omega^{455}\}, \{\omega^{107}, \omega^{180}, \omega^{253}\}\}.$$

- $\mathcal{C}_{\mathrm{Loc}} = \left\{(G_{\mathbf{m}}(\gamma), \gamma \in P) \mid \mathbf{m} \in \mathbb{F}_{2^9}^4\right\}$, and the local codes are $\mathcal{C}_j = \left\{(G_{\mathbf{m}}(\gamma), \gamma \in P_j) \mid \mathbf{m} \in \mathbb{F}_{2^9}^4\right\}$ for $1 \leqslant j \leqslant 3$

Rank-Distance Optimality of the Proposed Construction

Theorem: The proposed construction is Singleton-optimal, *i.e.*,

$$d_R\left(\mathcal{C}_{\mathrm{Loc}}\right) = n - k + 1 - \left(\left\lceil \frac{k}{r} \right\rceil - 1\right)(\delta - 1).$$

Proof Idea:

The proposed code $\mathcal{C}_{\mathrm{Loc}}$ is a subcode of an $\left(n, k + \left(\frac{k}{r} - 1\right)(\delta - 1)\right)$ Gabidulin code

- Example:
  - Recall our example, $n = 9, k = 4, r = 2, \delta = 2$
  - $G_m(x) = m_0 x^{2^0} + m_1 x^{2^1} + m_3 x^{2^3} + m_4 x^{2^4}$
  - This is a subcode of a $(9, 5)$ Gabidulin code, $d_R\left(\mathcal{C}_{\mathrm{Loc}}\right) = 5$

Rank-Locality of the Proposed Construction

Theorem: The proposed construction has (r, δ) rank-locality.

Proof Sketch:

- We write the encoding polynomial $G_{\mathbf{m}}(x)$ in terms of a good polynomial $H(x) := x^{q^{r+\delta-1}-1}$ as $G_{\mathbf{m}}(x) = \sum_{i=0}^{r-1} G_i(x) x^{q^i}$, where $G_i(x) = m_{i0} + \sum_{j=1}^{\frac{k}{r}-1} m_{ij} [H(x)]^{\sum_{l=0}^{j-1} q^{(r+\delta-1)l+i}}$.
- Define the repair polynomial for a $\gamma \in P_j$ as

$$R_j(x) = \sum_{i=0}^{r-1} G_i(\gamma) x^{q^i}.$$

- We show that $H(x)$ is constant on P_j, and thus, the evaluations of the encoding polynomial $G_{\mathbf{m}}(x)$ and the repair polynomial $R_j(x)$ on points in P_j are identical

Proposed Construction: Example

$n = 9, k = 4, r = 2, \delta = 2$. Set $q = 2$ and $m = n$

ω: primitive element of $\mathbb{F}_{2^9}$

- Encoding polynomial:

$$G_{\mathbf{m}}(x) = m_{00}x^{2^0} + m_{01}x^{2^3} + m_{10}x^{2^1} + m_{11}x^{2^4}$$

- Evaluation points:

  $P = \{P_1 = \{1, \omega^{73}, \omega^{146}\}, P_2 = \{\omega^{309}, \omega^{382}, \omega^{455}\}, P_3 = \{\omega^{107}, \omega^{180}, \omega^{253}\}\}$

- Repair polynomials:

$$R_1(x) = (m_{00} + m_{01})x^{2^0} + (m_{10} + m_{11})x^{2^1},$$
$$R_2(x) = (m_{00} + \omega^{119}m_{01})x^{2^0} + (m_{10} + \omega^{238}m_{11})x^{2^1},$$
$$R_3(x) = (m_{00} + \omega^{238}m_{01})x^{2^0} + (m_{10} + \omega^{476}m_{11})x^{2^1}$$

$\mathcal{C}_j$ can be obtained by evaluating the repair polynomials $R_j(x)$ on P_j

Subspace Codes [Koetter-Kschischang '08]

$\mathcal{P}_q(M)$: set of all subspaces of $\mathbb{F}_q^M$

$\mathcal{G}_q(M, n)$: set of all n-dimensional subspaces of $\mathbb{F}_q^M$

- A subspace code $\mathcal{C}$ is a non-empty subset of $\mathcal{P}_q(M)$ endowed with subspace metric

$$d_S(U, V) = \dim(U) + \dim(V) - 2\dim(U \cap V)$$

- The minimum subspace distance of a subspace code $\Omega \subseteq \mathcal{P}_q(M)$ is defined as

$$d_S(\Omega) = \min_{V_i, V_j \in \Omega,\, V_i \neq V_j} d_S(V_i, V_j)$$

- Constant-dimension code: A subspace code Ω in which each codeword has the same dimension, say n, *i.e.*, $\Omega \subseteq \mathcal{G}_q(M, n)$
- Such a code with minimum subspace distance d is denoted as an $(M, n, \log_q|\Omega|, d)$ code

(r, δ) Subspace-Locality

$[U]$: a matrix in a reduced column echelon form (RCEF) such that its columns span subspace U

$[U]\,|_S$: the sub-matrix of $[U]$ formed by columns indexed by $S \subset [n]$

$U\,|_S$: column space of $[U]\,|_S$

$\Omega\,|_S = \{U\,|_S : U \in \Omega\}$

- A constant-dimension subspace code $\Omega \subseteq \mathcal{G}_q(M, n)$ is said to have (r, δ) subspace-locality if, for each $i \in [n]$, there exists a set $\Gamma(i) \subset [n]$ such that
  1. $i \in \Gamma(i)$,
  2. $|\Gamma(i)| \leqslant r + \delta - 1$,
  3. $\dim\left(\Omega\,|_{\Gamma(i)}\right) = |\Gamma(i)|$, and
  4. $d_S\left(\Omega\,|_{\Gamma(i)}\right) \geqslant \delta$.
- The code $\mathcal{C}\,|_{\Gamma(i)}$ is said to be the local code associated with the i-th column

Lifting Construction [Silva-Koetter-Kschschang '09]

- X: codeword of a rank-metric code $\rightarrow \Lambda(X)$: subspace

$$\Lambda(X) = \left\langle \begin{bmatrix} I \\ X \end{bmatrix} \right\rangle,$$

  where I is $n \times n$ identity matrix, and $\langle . \rangle$ denotes the column space of a matrix

- $\Lambda(\mathcal{C}) = \{\Lambda(X) : X \in \mathcal{C}\}$: lifting of $\mathcal{C}$
- The subspace code constructed by lifting inherits the distance properties of its underlying rank-metric code

$$d_S(\Lambda(\mathcal{C})) = 2\, d_R(\mathcal{C})$$

Lifting Construction [Silva-Koetter-Kschschang '09]

- X: codeword of a rank-metric code $\rightarrow \Lambda(X)$: subspace

$$\Lambda(X) = \left\langle \begin{bmatrix} I \\ X \end{bmatrix} \right\rangle,$$

  where I is $n \times n$ identity matrix, and $\langle . \rangle$ denotes the column space of a matrix
- $\Lambda(\mathcal{C}) = \{\Lambda(X) : X \in \mathcal{C}\}$: lifting of $\mathcal{C}$
- The subspace code constructed by lifting inherits the distance properties of its underlying rank-metric code

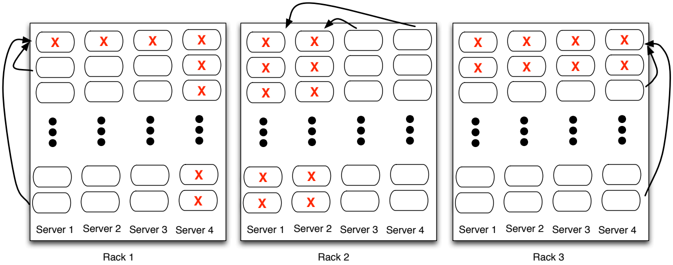
$$d_S\left(\Lambda(\mathcal{C})\right) = 2\, d_R\left(\mathcal{C}\right)$$

Theorem: Let $\mathcal{C}_{\mathrm{Loc}}$ be an $(m \times n, k, d, r, \delta)$ rank-metric code. The code $\Lambda(\mathcal{C}_{\mathrm{Loc}})$ obtained by lifting $\mathcal{C}_{\mathrm{Loc}}$ is an $(m+n, n, mk, 2d, r, 2\delta)$ subspace code.

Erasure Correction Capability

Theorem: A rank-metric code with (r, δ) rank-locality is guaranteed to locally correct the erasures and errors $E(C_j)$ and $E'(C_j)$ in a local array C_j provided $2\,\mathrm{rank}\,(E'(C_j)) + \mathrm{wt}_c\,(E(C_j)) \leqslant \delta - 1$.

- Follows from the rank-distance guarantee of a local code



Rank-metric code with $(2, 3)$ rank-locality can locally recover from crisscross erasures affecting any two rows and/or columns

Conclusion and Future Directions

- Rank-locality: Local codes possess good rank distance
  We computed tight upper bound on the rank-distance of codes with rank-locality and constructed optimal codes
- Subspace-locality: Local codes possess good subspace distance
  We obtained a class of subspace codes by lifting the proposed local rank-metric codes

Future Directions

- Can we construct rank-metric codes such that every column as well as row is associated with a local code?
- Can we improve the recovery performance by combining rank-metric decoding and Hamming-metric decoding for individual node failures?
- Can we investigate the impact of subspace-locality for repair over erroneous networks?