Lower Bounds for Locally Recoverable Codes

Itzhak (Zac) Tamo  
_Tel Aviv University_

Alexander Barg  
_University of Maryland_

Alexey Frolov  
_IITP, Russian Academy of Sciences_

_Coding Workshop - Technion_
Outline

• Locally Recoverable (LRC) Codes
• Lower bound for LRC codes with large alphabet size
• Lower bound for LRC codes with constant alphabet size (GV-type bound)
• Summary
Locally Recoverable Codes (LRC)

- \((n, k, r)_{LRC}\):
  - Takes \(k\) blocks produces \(n\) blocks
  - Any block has a recovering set of \(r\) other blocks, \(r \ll k\)
  - Clearly \(1 \leq r \leq k\)
Related Work

- *On the Locality of Codeword symbols* - Gopalan, Huang, Simitci, and Yekhanin

- Optimal Locally Repairable Codes - Rawat, Koyluoglu, Silberstein, Vishwanath,

- Optimal Locally Repairable Codes – T., Papaliopoulos, Dimkis

- Partial MDS codes - Blaum, Plank, Schwartz, Yaakobi

- Vijay Kumar’s group, Camilla Hollanti’s group, and Paul Siegel’s group

- And many more...
Locality and Minimum Distance

- \((n, k, r)\)LRC:
  - \(d \leq n - k + 2 - \lceil k/r \rceil\)  \(\text{Gopalan, Huang, Simitci, and Yekhanin}\)
  
  - Observation: Smaller locality \(\rightarrow\) lower failure resilience

  - Optimal \((n, k, r)\)LRC has minimum distance \(d = n - k + 2 - \lceil k/r \rceil\)

  - There exists an optimal \((n, k, r)\) LRC over small finite fields \(\text{T. and Barg}\)
    - \(|F| = n\)
    - Generalization of RS Codes
LRC Codes and Multiple Recovering Sets

- $(n, k, r, t)$ LRC:
  - Any symbol has $t$ disjoint recovering sets of size $r$
  - $d \leq n - \sum_{i=0}^{t} \left\lceil \frac{k-1}{r^i} \right\rceil$

  T. and Barg
  Rawat, papaiopoulos, Dimakis and Vishwanath

- $t = 0$: $d \leq n - k + 1$ Tight bound

- $t = 1$: $d \leq n - k + 2 - \left\lceil k/r \right\rceil$ Tight bound

- $t = 2$: $n - k - \left\lfloor \frac{2k}{r-1} \right\rfloor + 1 \leq d \leq n - (k - 1 + \left\lceil \frac{k-1}{r} \right\rceil + \left\lceil \frac{k-1}{r^2} \right\rceil)$

- $t = 2$: $1 - R - \frac{2R}{r-1} \leq \delta \leq 1 - R - \frac{R}{r} - \frac{R}{r^2}$
LRC Codes and Multiple Recovering Sets

- **Theorem:** Let $R \leq 1 - \frac{t}{r+1}$ and $\delta$ that satisfies

  $$\delta = \frac{1 - t - R}{1 - t \gamma}$$

  $$\frac{t-1}{t} h(\delta) - \frac{1}{r+1} h(\delta \gamma (r + 1)) - \delta \gamma (r + 1) h \left( \frac{1}{\gamma (r + 1)} \right) = 0$$

  then for sufficiently large alphabet there exists an $(R, \delta, r, t)$ LRC

- **Remarks:**

  - Such $0 \leq \delta \leq 1$ always exists
  - This bound is better than all existing lower bounds
  - The proof uses the existence of graphs with good expansion properties (expander graphs)
Some Plots

\[ r = 10, t = 3 \]

\[ r = 9, t = 2 \]
From Expander Graphs to LRC Codes with Multiple Recovering Sets

- $G = (V = V_1 \cup V_2, E)$ a biregular graph:
  
  - $|V_1| = n$ and $\deg(v) = t$ for $v \in V_1$
  
  - $|V_2| = \frac{nt}{r+1}$ and $\deg(v) = r + 1$ for $v \in V_2$

  - $G$ is $(\alpha, \gamma)$ -- expander if every subset $T \subset V_1, |T| \leq \alpha n$ has at least $\gamma t |T|$ neighbors in $V_2$
From Expander Graphs to LRC Codes with Multiple Recovering Sets

- The parity check matrix of the code:

\[ H = \begin{bmatrix} A_{\frac{nt}{r+1} \times n} & \text{adjacency matrix of the graph} \\ H_{n-k-\frac{nt}{r+1} \times n} & \text{matrix picked randomly} \end{bmatrix} \]

- The nonzero entries of the adjacency matrix \( A \) are picked uniformly at random from the field \( F^* \).
Why Expansion?

- Need to show that every $\delta n$ columns are linearly independent.

$$H =$$

- Lemma: if the $\delta n$ columns "touch" at least $\delta n$ rows then w.h.p the columns are linearly independent.

$A_{\frac{nt}{r+1} \times n}$ adjacency matrix of the graph

$H_{n-k-\frac{nt}{r+1} \times n}$ matrix picked randomly
Disjoint Recovering Sets

- Disjoint recovering sets $\leftrightarrow$ The girth of $G$ is greater than 4

- Theorem (Bollobas): The number of cycles of length $l$ in a uniformly picked regular graph behaves like a poisson r.v.

- Corollary: $P(\text{girth of } G > 4) > \varepsilon$ for some $\varepsilon > 0$

- Theorem (Burstein and Miller): $\lim_{n \to \infty} P(G \text{ has } "\text{good}" \text{ expansion}) = 1$

- Corollary: For large enough $n$ there exists $G$ with good expansion and girth $> 4$
GV-Type Bound for Constant Alphabet
Related Work

- Binary Cyclic Codes that are Locally Repairable - Goparaju and Calderbank (ISIT 2014)
- Optimal Linear and Cyclic LRC Codes over Small Fields – Zeh and Yaakobi (ITW 15)
- Achieving Arbitrary Locality and Availability in Binary Codes – Wang and Zhang
- An Upper Bound On the Size of LRC Codes – Cadambe and Mazumdar

  - Alphabet dependent bound on \((n, k, r)_q\) LRC Code

\[
k \leq \min_{1 \leq m \leq \min \left( \left\lfloor \frac{n}{r+1} \right\rfloor, \left\lceil \frac{k}{r} \right\rceil \right)} \left\{ tr + k_q(n - m(r + 1), d) \right\}
\]
For $t = 1$ recovering set, the following asymptotic GV-type bound holds true:

1) \[ R_q(r, \delta) \geq \frac{r}{r+1} - \min_{0 < s \leq 1} \left\{ \frac{\log_q b(s)}{r+1} - \delta \log_q s \right\} \]

2) \[ b(s) = \frac{1}{q} \left( (1 + (q - 1)s)^{r+1} + (q - 1)(1 - s)^{r+1} \right) \]
GV-Type Bound (Constant Alphabet)

- Construct an \((n - k) \times n\) parity check matrix

\[
H = \begin{bmatrix} H_U \\ H_L \end{bmatrix}, \quad H_U = \begin{bmatrix} H_0 \\ & \ddots \\ & & H_0 \end{bmatrix}
\]

- \(H_U\) has \(\frac{n}{r+1}\) copies of \(H_0\)

- \(H_0\) is a simple parity check code of length \(r + 1\)

- \(H_L\) is an \(\left(\frac{nr}{r+1} - k\right) \times n\) matrix with entries picked uniformly and independently from \(F_q\)
GV-Type Bound (Constant Alphabet)

- \( H \cdot x = 0 \iff H_U \cdot x = 0 \) and \( H_L \cdot x = 0 \)

- Upper bound the number of “bad” (low weight) vectors that satisfy \( H_U \cdot x = 0 \) using the weight enumerator \( b(s) \) of the code checked by \( H_U \)

- \( H_U = \begin{bmatrix} H_0 & \cdot & \cdot & \cdot & H_0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & H_0 \\ \end{bmatrix} \)

- The weight enumerator of the code generated by \( H_U \) is easy to derive

- Using MacWilliams Identity derive the weight enumerator of the code with parity check matrix \( H_U \):

\[
b(s) = \frac{1}{q}((1 + (q - 1)s)^{r+1} + (q - 1)(1 - s)^{r+1})
\]
GV-Type Bound (Constant Alphabet)

- Let $M$ be the upper bound on the bad (low weight) vectors

$$M \geq |\{x: H_U \cdot x = 0, \text{wt}(x) < \delta n\}|$$

- The entries of $H_L$ are picked randomly →
  
  for $x \neq 0$ then $P(H_L \cdot x = 0) = q^{-n\left(\frac{r}{r+1}-R\right)}$

- If $M \cdot q^{-n\left(\frac{r}{r+1}-R\right)} < 1$ then there exists a parity check matrix $H = \begin{bmatrix} H_U \\ H_L \end{bmatrix}$ with minimum distance at least $\delta n$ (union bound)
For $t = 2$ recovering sets, the following asymptotic GV-type bound holds true:

1) $R_q^{(2)}(r, \delta) \geq \frac{r}{r + 2} - \min_{0 < s \leq 1} \left\{ \frac{1}{\binom{r+2}{2}} \log_q g^{(2)}_q(s) - \delta \log_q s \right\}$

2) $g^{(2)}_2(s) = \frac{1}{2r+2} \sum_{i=0}^{r+2} \binom{r+2}{i} (1 + s)^{r+2-i} (r+2-i) (1 - s)^{i(r+2-i)}$. 

GV-Type Bound (Constant Alphabet) - 2 Recovering sets
GV-Type Bound (Constant Alphabet)  
2 Recovering sets

(i) Singleton Bound

(ii) GV - bound on $R_2^{(2)}(r, \delta)$

(iii) GV - bound on $R_2(r, \delta)$

(iv) Plotkin bound

(b) Asymptotic bounds on codes with one and two recovering sets.
GV-Type Bound (Constant Alphabet)

• Construct an \((n - k) \times n\) parity check matrix

\[
H = \begin{bmatrix}
H_U \\
H_L
\end{bmatrix},
H_U = \begin{bmatrix}
H_0 \\
\ddots \\
H_0
\end{bmatrix}
\]

• \(H_U\) has \(\frac{n}{r+2}\) copies of \(H_0\)

• \(H_0\) is an \((r + 1) \times \binom{r + 2}{2}\) edge vertex incidence matrix of a complete graph \(K_{r+2}\) with one row deleted.

• \(H_L\) is an \(\binom{nr}{r+2} - k \times n\) matrix with entries picked uniformly and independently from \(F_q\)
Summary

- Lower bounds for \((n, k, r, t)\) LRC code over large alphabet using expanders
  - Derive tighter Singleton-type bounds
  - Derandomize the construction using
    - Known expander graphs
    - Explicit assignment to the entries of the parity check matrix
- Lower bounds for \((n, k, r, t = 1,2)\) LRC code with constant alphabet size (GV-type bound)

Thank you for listening