LECTURE 19

TODAY + Next 2 lectures

LOCALITY IN CODING THEORY

- (Today) LOCAL RECONSTRUCTION CODES (LRC)
- (Next) LOCALLY DECODABLE CODES (LDC)
- (Later) LOCALLY TESTABLE CODES (LTC)

"LOCALITY": Algorithms that work without reading the whole input. (Instead sample input).

"SUBLINEAR TIME ALG"

Local Recovery:

- Very practically motivated (cloud servers)
- two concerns
  - : mild failure : 1 erasure; Need to "patch" quickly
  - : catastrophic failure : many erasures, just need large distance.

Contrast with

local Decoding: constant fraction errors!

more of theoretical interest; decoding randomized

·(necessarily so!).

ℓ-LRC: $(n,k,d)_q$-code C is an ℓ-LRC

if $\forall\, i \in [n]\ \exists\, S_i \subseteq [n]\setminus\{i\}$, $|S_i| \le \ell$

s.t. $S_i \Longrightarrow_C \{i\}$ if $T \subseteq S$ $\forall\, C$

$(S \Longrightarrow_C T)$ if $x, y \in C$ $S \Longrightarrow_C T$

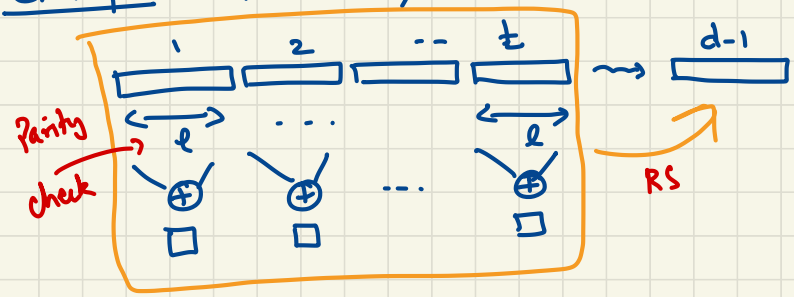
"S determines T in C" & $x|_S = y|_S \Rightarrow x|_T = y|_T$

ℓ-mLRC [message LRC] if C is systematic $m \mapsto (m, c)$

& $\forall\, i \in [k]\ \exists\, S \subseteq [n]\setminus\{i\}$ s.t. $S \Longrightarrow_C \{i\}$, $|S| \le \ell$

C ℓ-LRC $\Rightarrow$ C ℓ-mLRC

Example: $k = \ell \cdot t$; wanted dist. d



Assume large alphabet:

$$m = m_1, m_2 \dots m_t \quad ; \quad m_i \in \mathbb{F}_q^{\ell}$$

$$x_i \triangleq \sum_j m_{ij}$$

$$m \longmapsto (m, x_1 .. x_t) \xrightarrow{RS} RS(m, \bar{x})$$

$$\mathbb{F}_q^{k} \qquad \mathbb{F}_q^{k+t} \qquad \mathbb{F}_q^{k+t+d-1}$$

Achieves $n = k + \frac{k}{\ell} + d - 1$

Thm: if C is an $(n,k,d)_q$ code & an ℓ-MLRC

then

$$n \geq k + \frac{k-1}{\ell} + d - 1$$

Proof: Let C be a systematic code.

- **Idea:** find $U \subseteq [n]$ with $|U| = k-1$
  & $|V| \subseteq [n]$ with $|V| = \left(\frac{k-1}{\ell}\right) + k-1$
  $= \frac{(k-1)(\ell+1)}{\ell}$

$\forall i$ $S_i \Rightarrow S_i \cup \{i\}$ s.t. $U \underset{\ell}{\Longrightarrow} V$

$\Rightarrow n \geq \frac{k-1}{\ell} + k-1 + \delta$

Bound follows. (Why?)

- finding U, V: iterate
  - initially $U = \phi$; $V = \phi$
  - iteratively $U \leftarrow U + (S_i \setminus V)$; $V \leftarrow V + S_i + i$
  - repeat till $|V| - |U| = \left(\frac{k-1}{\ell}\right)$;
  - then add arbitrarily to U & V.

PHP $\Rightarrow \exists\, x \neq y \in C$ s.t. $x_U = y_U$
$\Rightarrow x_V = y_V$
$\Rightarrow \delta(x,y) \leq n - |V|$

---

① invariants $U \underset{\ell}{\Rightarrow} V$ always; $U \subseteq V$

② $|V \cap [k]| \leq k-1$; $|U| \leq k-1$

(Why?)

Thm [Tamo-Barg]: $\exists$ ℓ-LRC C that $[n,k,d]_q$

with $n \leq k+d-1+\frac{k-1}{\ell}$

- LRC not mLRC !
- But matches lower bound for mLRC.

Main ideas in construction:

1. Subcode of Reed-Solomon code.
2. Choose field carefully.
3. Clever algebra in analysis

Field: $q = (\ell+1)\cdot t + 1$ $\Leftarrow$ $\ell+1 \mid q-1$, $q \geq$

$\Rightarrow \exists \alpha \in \mathbb{F}_q$ s.t. $\alpha^{\ell+1} = 1$; $\alpha^i \neq 1$ for $i \leq \ell$.

$\forall \beta \in \mathbb{F}_q^*$ $\beta^{q-1} = 1$; $\alpha^{\ell+1} = 1$

$\Rightarrow (\alpha^{(\ell+1)t} = 1)$

Code: $n \triangleq q-1 \qquad r \triangleq \ell+1$

$$M \triangleq \left\{ i \mid 0 \le i < \frac{(\ell+1)\cdot k}{\ell},\ i \neq -1 \bmod r \right\}$$

$$C = \left\{ f : \mathbb{F}_q^* \to \mathbb{F}_q \;\middle|\; f = \sum_{i \in M} c_i x^i \right\}$$

$x^0, x^1 \dots x^{\ell-1}, 0, x^{\ell+1}, x^{\ell+2} \dots 0$ ($\ell+1$)

Parameters:

- $n = q-1$
- $\dim = k$ (assume $\ell \mid k$)
- $\text{dist} = n - \frac{(\ell+1)k}{\ell} + 1$
- locality = ?

$|M| = \dim = k$

Locality: $\forall\, a \in \mathbb{F}_2^*$ let $S_a \triangleq \{a, \alpha a, \alpha^2 a, \ldots \alpha^l a\}$

- Claim: for every $f \in C$ $\left.\begin{array}{l} \\ \\ \end{array}\right\} S_a - a \Rightarrow_C a$
  $f|_{S_a}$ are dependendent!
- Proof:
  - $f|_a \equiv f \bmod (x-a)$
  - $f|_S \equiv f \bmod \prod_{a \in S} (x-a)$
  - for S_a:

Exercise: Put floors & ceiling in all results.

$$\prod_{a \in S} (x-a) = x^r - a^r$$ ← Exercise

- $\Rightarrow f|_{S_a} = f(x) \bmod (x^r - a^r) = g_a(x)$

$\sum_{i=0}^{l} f(a\alpha^i) = 0$ $\forall f \in C$ — Exercise

$\deg g_a = ?$ trivially $\le r-1$

But we skipped all i s.t. $i = -1 \bmod r$

$\Rightarrow \deg g_a \le r-2$

$\Rightarrow$ values at $|S_a| = r$ points constrained!

$\Rightarrow g_a(a\alpha), \ldots g_a(a\alpha^l) \Rightarrow g_a(a)$

$\Rightarrow f(a\alpha) \ldots f(a\alpha^l) \Rightarrow f(a)!$

$f(x) = x^7 + 3x^6 + 5x^4 - 2x^3 + 3x^2 + x + 1$

$\mod (x^3 - 1) = (x^3 - x^0)$

$1 + x + 3x^2$

$+ \quad -2x^{3-0} + 5x^{4-1} + 0 \cdot x^{5-2}$

$+ \quad 3x^{6-0} + x^{2-1}$

$2 \cdot x^0 + 7x + 3x^2$

$f : D \to R$

$S \subseteq D$

$f|_S : S \to R$

$S \underset{C}{\Longrightarrow} T$

is the function $f|_S(a) = f(a)$

$\forall f \in C$

$f|_S \Rightarrow f|_T$

Preview of next two lectures

- Locally decodable & testable codes
- Defn. by examples:

(1) Hadamard Code $H_n = \left\{ f : \mathbb{F}_2^n \to \mathbb{F}_2 \;\middle|\; f = \sum \alpha_i x_i \right\}$

- Claims: - H_n is $\left(2, \frac{1}{4}\right)$-locally-decodable.

$\forall\, f \in H_n, g, x$ if $\delta(f,g) < \frac{1}{4}$

then $f(x) = \underset{y}{\text{majority}} \{ g(x+y) - g(y) \}$

- H_n is $\left(O\left(\frac{1}{\epsilon^2}\right), \frac{1}{2} - \epsilon\right)$-locally-list-decodable.

Famed "Goldreich-Levin" algorithm.

roughly can "list-decode" from $\frac{1}{2} - \epsilon$ fraction error "locally" with $O(1/\epsilon^2)$ queries

- H_n is $(3, 1)$-locally-testable.

Famed Blum-Luby-Rubinfeld linearity test.

Thm: $\forall g \; \exists f \in H_n$ s.t.

$$\delta(f,g) \leq 1 - \Pr_{x,y}\left[g(x) + g(y) \neq g(x+y) \right]$$

$$f(x) = x^7 + 3x^6 + 5x^4 - 2x^3 + 3x^2 + x + 1$$

$$\text{mod}\ (x^3 - 1) = (x^3 - x^0)$$

$$= 1 + x + 3x^2$$

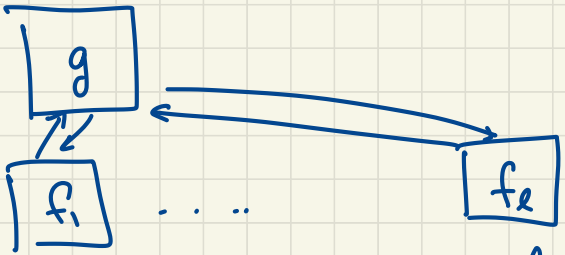
$$S_a = \sum \alpha^i \cdot f(\alpha^i a) = 0$$

$$\sum \alpha^0 \neq 0 \qquad \sum \alpha^i = 0 \quad \forall\, i \neq 0$$

Goldreich-Levin

Local list-decoding

given box

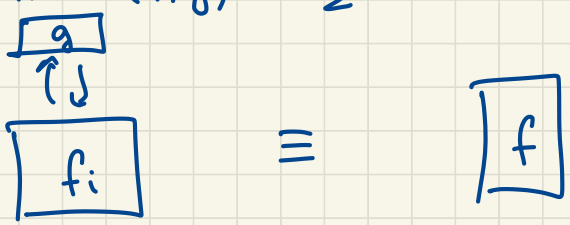


construct boxes

$\ell = O(\frac{1}{\epsilon^2})$

s.t $\forall f \in H_n \quad \delta(f,g) \le \frac{1}{2} - \epsilon \quad \exists i$

s.t.



Local-list-decoder for H_n

- $t = \log \frac{1}{\epsilon^2}$; Pick $a_1 \dots a_t \in \mathbb{F}_2^n$ randomly

$f_{b_1 \dots b_t}(x) \triangleq$ Let $L_0(y_1 \dots y_t, y_{t+1}) = \sum_{i=1}^{t} b_i y_i$

$L_1(y_1 \dots y_t\, y_{t+1}) = \sum b_i y_i + y_{t+1}$

$A = \{\sum y_i a_i + x \mid y_1 \dots y_t\}$

Of L_0 & L_1 pick one that has larger agreement with g on A.

Next two lectures

(1) Higher degree Polynomials (1th = linear poly)

(2) Best "3-query LDC"

Decoding + testing will $n^{o(1)}$-queries close to Singleton Bound!