Today:
Introduction to locality in coding
- Definitions of local decodability
  local testability
  local recoverability
- Reed-Muller Example
- "Optimal" LRCs.

Locality in Algorithms
- Algorithms usually compute functions
  \[ f: \{0,1\}^n \to \{0,1\}^m \]
- Running time always \( \Omega(n+m) \) - right?
- Can we conceive algorithms running in time \( o(n+m) \)?
- What expectations/guarantees can we have?
- Are they useful?
Will illustrate in context of decoding.
Decoding Algorithm

- Let $C \subseteq \{0,1\}^n$ have encoding function $E: \{0,1\}^k \rightarrow \{0,1\}^n$

- Decoder should map $D: \{0,1\}^n \rightarrow \{0,1\}^k$

- But suppose we want only $m_i$ for message $m \in \{0,1\}^k$, given $y \approx E(m)$.

- Is it essential to read all of $y_1, \ldots, y_n$? Or can we get away by "sampling"?

- Answer: Not obvious....

  In retrospect: YES.

- $E$ locally decodable

- Local (decoding) algorithm: $A_y$
  . Picks distribution $q$ over $\binom{[n]}{l}$
  . Samples $s \sim P$
  . Queries $y_s \subseteq \{y_i; 1 \leq i \leq n\}$
  . Answer based (only) on $y_s$. [Independent of $y_s$]
Example: Reed-Muller Codes

Recall: Reed-Muller Codes $\text{RM}(q, m, d)$

\[ \text{RM}(q, m, d) \triangleq \{ f : \mathbb{F}_q^m \to \mathbb{F}_q \mid \deg(f) \leq d \} \]

Field size $\Rightarrow$ Variables $\Rightarrow$ degree.

represented as vector of length $n = q^m$ over $\mathbb{F}_q$.

\[
\begin{align*}
\text{distance of } \text{RM}(q, m, d) & \Rightarrow q^m \cdot \frac{(q^d-q)}{q^m} = q^{m-1}(q-d) \\
\text{dimension} & = \binom{d+m}{m} \geq \exp(\min(d,m)) \\
n = q^m; & \quad k \geq \exp(\min(d,m)) \\
\text{But will get } l \leq q! & \quad (\text{even } l = d+2)
\end{align*}
\]

Main Idea: "Local Redundancies":

Line $L_{a,b} \triangleq \{ a + t \cdot b \mid t \in \mathbb{F}_q^m \}$ (a, b $\in \mathbb{F}_q^m$)

\[ f_{L_{a,b}}(t) \triangleq f(a + t \cdot b) \quad [ f_{L_{a,b}} \in \mathbb{F}_q^2 ] \]

\[ \deg(f_{L_{a,b}}) \leq \deg(f) \quad ; \quad d < q-1 \Rightarrow f_{L_{a,b}} \text{ is not an arbitrary function } \mathbb{F}_q \to \mathbb{F}_q \]
$(x, y)$ local decoder: $D_y(i), \ i \in [k]$ outputs $m_i$ w.p. $\geq \frac{2}{3}$

if $S(E(m), y) \leq \epsilon \cdot S(C)$

distance of code $C$.

$C$ is $(x, \epsilon)$-locally decodable if it has an $(x, \epsilon)$-local decoder $D$

$(x, \epsilon)$-local tester: $Ty$

Outputs YES if $y \in C$.

Outputs NO w.p. $\geq \epsilon \cdot S(y, C)$.
Linearity of $RM(q, m, d)$ + Local Redundancy

\[ \Rightarrow \exists x_0 \in \mathbb{F}_2^m \land x_1 \ldots x_e \in \mathbb{F}_2^m \text{ s.t.} \]

\[ f(x_0) \text{ determined by } f(x_1) \ldots f(x_e) \]

- Symmetry: \[ \Rightarrow x_0 = 0 \; ; \; \{ x_1 \ldots x_e \} = \mathbb{F}_2 \setminus \{0\} \text{ work.} \]

\[ f|_L(0) = \sum_{i=1}^{q-1} x_i \cdot f|_L(x_i) \quad \text{for } x_1 \ldots x_{q-1} \in \mathbb{F}_2 \setminus \{0\} \]

Local Decoding Problem: given oracle access to \( f : \mathbb{F}_2^m \rightarrow \mathbb{F}_2 \)

\[ \exists \text{ poly } p : \mathbb{F}_2^m \rightarrow \mathbb{F}_2 \text{ s.t. } S(f, p) \leq ?? \]

- a \in \mathbb{F}_2^m \; ; \; \text{compute } f(a)

Local Decoder: \( D(a) : \)

- Pick \( b \in \mathbb{F}_2^m \) at random; let \( \xi = \frac{f(a) + \eta b}{2} \).

- Output \[ \sum_{i=1}^{q-1} \xi_i \cdot f(a + \eta_i b) \]

makes \( q-1 \) queries \( \Rightarrow (q-1) \)-local.

Correctness? \( \epsilon = ? \)
Analysis:

1. Least for random $b$ & $\eta_i \neq 0$

   $a + b\eta_i$ is random independent of $a$.

   $\implies \Pr_b [ f(a + b\eta_i) = p(a + b\eta_i) ] \leq S(f, p)$

2. $\implies \Pr_b [ \exists i \in \{1, 2, \ldots, n\} : f(a + b\eta_i) = p(a + b\eta_i) ] \leq (q-1) \cdot S(f, p)$

   if $\forall i : f(a + b\eta_i) = p(ab\eta_i)$ then

   $\sum_{i=1}^n f(ab\eta_i) = \sum_{i=1}^n p(ab\eta_i) = p(a)$

   $\implies$ if $(q-1) \cdot S(f, p) \leq \frac{1}{3}$ then $D_f(a) = p(a)$ w.p. $\geq \frac{2}{3}$

Yields $\varepsilon = \frac{1}{3(q-1)}$

Food for thought: Improve $\varepsilon$ to $\varepsilon(1)$.

(Exercise): Reduce $\varepsilon$ to $\varepsilon(0)$. 

Local testability:

Test: Verify the $\deg(f|_L) \leq d$ for random $h = \ell_{a,b}$

Locality: $\ell = q$

Analysis: Non-trivial

Obstacles: if even if $\exists L \; \deg(f|_L) \leq d \leq q$

if maybe that $\deg(f) > d$

Example: $q = 2^\ell; \; d = 2^{\ell-1}; \; f(x_1, \ldots, x_m) = x_1^d \cdot x_2^d$

on line $L_{a,b}$: $x_1 = a_1 t + b_1$

$x_2 = a_2 t + b_2$

$f(x_1, \ldots, x_m) = (a_1 t + b_1)^d \cdot (a_2 t + b_2)^d$

$= (a_1 a_2)^d \cdot t^{2d} + (a_1 b_2 + a_2 b_1)^d \cdot t^d + (b_1 b_2)^d$

$= (a_1 a_2)^d \cdot t^d + (a_1 b_2 + a_2 b_1)^d \cdot t^d + (b_1 b_2)^d$

This implies $\deg(f|_L) \leq \ell \cdot d$.

(under various conditions)

Nevertheless: Theorem: $\Pr(\deg(f|_L) \leq d) \leq \epsilon \Rightarrow S(f, \ell_m) \leq 2\epsilon$. 

General Questions:

Low Query Regime:

1. What is best relationship between $[n, k, d]_2$ if we want code to $(\epsilon, \eta)$-LDC with $\epsilon = 2(\frac{1}{e})^l$ and $l \leq 2, 3, 4$ ... 
   \[d = 2(n^l); \quad n \geq k^{1+\frac{1}{e}}; \quad n \leq \exp(\exp(\frac{1}{\log(k)})\]

2. What is best $[n, k, d]_2$ if we want code to be $(\epsilon, \eta)$-LTC with $\epsilon = 2(\frac{1}{e})^l$ and $l \leq 2, 3, 4$ ...
   \[n = O(k \text{ polylog } k) \text{ with } l = 3.

High Rate Regime

If we want $R \approx 1-\delta$ what is the smallest $l$ we can achieve

- LDC: $l \leq 2^{\frac{1}{\log(n)}}$ ...
- LTC: $l \leq (\log(n))^{\log(n)}$

[Next two lectures: Ideas]
Rest of Today: $10^9$ $\text{ Aside}$

Practical Motivation: [Gopalan, Huang, Simitci, Yekhanin '12]

In Cloud Storage: two kinds of faults:

A - {1 Server/Memory bank goes down periodically;}
B - {Several servers may go down rarely;}

Want to recover from both; however
if A: then want very quick recovery ("local")
if B: slow recovery is O.K.
error model = erasure.

$LRC$ code: 1 distance $\geq d$ [Corrects d-1 erasures]

$\uparrow$
reconstructible?
$
\times$
reparable?

$LRC$ code $\left[ n \right] = \left[ k \right] \cup \left[ n-k \right]$

weak $\circ$
message $i \in \left[ k \right] \land \text{ codeword } C_i \ldots$

$C_i$ can be recovered from $C_S$ for some $|S| \leq L$ at $i$'s

strong $\bullet$
message $i \in \left[ n \right]$

$C_i$ can be recovered from $C_S$ for some $|S| \leq L$, at $i$'s
Question: What is the relationship between $n, k, e, d (q \rightarrow d)$?

[GIHY]:

[Trivial]: Weak recovery possible with

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

Given $m = m_1, m_2, m_3, \ldots, m_k$

$$+ \quad +$$

Encoding $= m \bigoplus \bigoplus_{i=1}^{k/e} m_{i(e-1)+i} + \sum_{j=1}^{k/e} RS^{+}(m)$

Parity check bit in systematic RS code of distance $d$.

Locality here

Distance here

[GIHY]: $n > k + \frac{k}{e} + d - O(1)$.

Proof:
- Find blocks of size $(e+1)$ of rank $\leq e$.
- Union them to get $(k-1) \left( \frac{e+1}{e} \right)$ coordinates $j$

$$\text{rank} \leq (k-1)$$

- Apply PHP to conclude

$$d \leq n - (k-1)(e+1) \text{ or } n \geq k + \frac{k}{e} + d$$
[Tam- Burg] \exists \text{ Strong } (k; d) \text{- codes with } n = k + \frac{k}{d} + d = O(1).

Construction: \[ q = (k+1)^{-1} \quad 2 = \frac{(k+1)}{s} + 1 \]

Such \( \mathbb{F}_q \) has \( w \) s.t. \( 1, w, w^2, \ldots, w^{y-1} \) distinct and \( w^y = 1 \).

Code: \[ \{ f(x) \mid x \in \mathbb{F}_q^* \} \quad f(x) = \sum a_i x^i \]

\[ B_{t+1} \frac{a_{t+1} \ldots a_1}{a_0} = 0 \]

Code: \[ \{ f(f(x)) \mid x \in \mathbb{F}_q^* \} \]

\[ C_k \{ \langle f(x) \rangle \mid x \in \mathbb{F}_q^* \} \quad f(x) = \sum a_i x^i \quad i \leq k \]

\[ \alpha_i = 0 \text{ if } i = -1 \mod y \]

\[ \alpha_{r-1}, \alpha_{2r-1}, \alpha_{3r-1}, \ldots = 0 \]

\[ \dim(C_k) = \frac{R}{y} \cdot \frac{(y-1)}{y} \]

\[ \text{dist}(C_k) = n - k \]

Locality?

Let \( S_a = \sum a, a \cdot w, a \cdot w^2, \ldots, a \cdot w^l \)

\[ f_1 = f(x) \left( \mod \prod (x-b) \right) = f(x) \mod (x^y - a^y) \]

\[ f_1 \mid_{S_a} \]
But $f(x) \mod (x^n - a^n)$ has degree $\leq n-2$!

[One smaller than it should be!]

so. $f(a)$ determined by $f|_{a^n - a^n} = a$.

Next Lectures

1. High Rate LDCCs
2. Some analysis of LTCs