3/30/17

CS 229r - LECTURE 18

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 concieve 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 $E : \{0,1\}^k \to \{0,1\}^n$
- Decoder should map $D : \{0,1\}^n \to \{0,1\}^k$
- 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.
- ~~[illegible] is (ϵ, δ) locally decodable~~
- ℓ-local (decoding) algorithm : A^y
  - Picks distribution P over $\binom{[n]}{\ell}$
  - Samples $S \sim P$;
  - Queries $y_S \triangleq \{y_i \mid i \in S\}$
  - Answers based (only) on y_S.
    [independent of $y_{\bar{S}}$]

Example: Reed-Muller Codes

Recall: Reed-Muller Codes $RM(q, m, d)$

Field size (q) #variables (m) degree (d).

$$RM(q,m,d) \triangleq \{ \underbrace{f : \mathbb{F}_q^m \to \mathbb{F}_q}_{\text{represented as vector of length } n = q^m \text{ over } \mathbb{F}_q} \mid \deg(f) \le d \}$$

$d<q$:

distance of $RM(q,m,d) = q^m \cdot \frac{(q-d)}{q} = q^{m-1}(q-d)$

dimension $= \binom{d+m}{m} \ge \exp(\min(d,m))$.

$n = q^m$; $k \ge \exp(\min(d,m))$

But will get $\ell \le q$! (even $\ell = d+2$)

Main Idea: "Local Redundancies":

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

$f|_{\ell_{a,b}}(t) \triangleq f(a + t \cdot b)$ $[f|_\ell \in \mathbb{F}_q^q]$

$\forall a,b$ $\deg(f|_\ell) \le \deg(f)$; $d < q-1 \Rightarrow f|_\ell$ is not an arbitrary function $\mathbb{F}_q \to \mathbb{F}_q$.

(ϵ, ℓ)- local decoder : $D^y(i)$ $\quad i \in [k]$

outputs m_i w.p. $\geq \frac{2}{3}$

if $\delta(E(m), y) \leq \epsilon \cdot \underbrace{\delta(C)}_{\text{distance of code } C.}$

— x —

C is (ϵ, ℓ)- locally decodable if it has an (ϵ, ℓ)- local decoder D

— x —

(ϵ, ℓ) - local tester : T^y

Outputs YES if $y \in C$.

Outputs NO w.p. $\geq \epsilon \cdot \delta(y, C)$

— x —

- Linearity of $RM(q, m, d)$ + local Redundancy

$\Rightarrow \exists\ x_0 \in \mathbb{F}_q^m$ & $x_1 \dots x_\ell \in \mathbb{F}_q^m$ s.t.

$f(x_0)$ determined by $f(x_1) \dots f(x_\ell)$

- Symmetry : $\Rightarrow x_0 = 0$; $\{x_1, \dots x_\ell\} = \mathbb{F}_q \setminus \{0\}$ work.

$$f|_L(0) = \sum_{i=1}^{q-1} \lambda_i f|_L(\eta_i) \qquad \{\eta_1 \dots \eta_{q-1}\} = \mathbb{F}_q \setminus \{0\}.$$

$[\exists \lambda_1 \dots \lambda_{q-1}$ s.t. $\dots]$

Local Decoding Problem: given oracle access to $f : \mathbb{F}_q^m \to \mathbb{F}_q$

s.t. $\exists$ deg d poly $p : \mathbb{F}_q^m \to \mathbb{F}_q$ s.t.

$\delta(f, p) \leq ???$

& $a \in \mathbb{F}_q^m$; compute $f(a)$

Local Decoder : $D^f(a)$:

- Pick $b \in \mathbb{F}_q^m$ at random ; ~~let $L = \ell_{a,b}(t)$~~.
- Output $\displaystyle\sum_{i=1}^{q-1} \lambda_i f(a + \eta_i b)$

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

Correctness ? $\epsilon = ?$

Analysis:

1. ~~Idea 1~~: for random b & $\overset{\text{fixed}}{\eta_i} \neq 0$

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

$$\Rightarrow \Pr_b \left[f(a + b \cdot \eta_i) \neq p(a + b \cdot \eta_i) \right] \leq \delta(f, p)$$

2. $$\Rightarrow \Pr_b \left[\exists i \in \{1 .. q-1\} \text{ s.t. } f(a + b \cdot \eta_i) \neq p(a + b \cdot \eta_i) \right] \leq (q-1) \cdot \delta(f, p)$$

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

$$\sum \lambda_i f(a + b \cdot \eta_i) = \sum \lambda_i \, p(a + b \cdot \eta_i) = p(a)$$

$$\Rightarrow \text{ if } \boxed{(q-1) \cdot \delta(f, p) \leq \tfrac{1}{3}} \text{ then } D^f(a) = p(a) \text{ w.p. } \geq \tfrac{2}{3}$$

yields $\epsilon = \dfrac{1}{3(q-1)}$.

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

(Exercise). Reduce ℓ to $O(d)$.

Local testability:

Test: Verify the $dg(f|_L) \le d$ for random $L = \ell_{a,b}$

locality: $\ell = q$

Analysis: Non-trivial

Obstacles: if even if $\forall L$ $deg(f|_L) \le d \le q$

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

$x_2 = a_2 t + b_2$

$$f(x_1 \ldots x_m)\Big|_L = (a_1 t + b_1)^d (a_2 t + b_2)^d$$

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

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

$$\boxed{t^{2d} = t^q = t}$$

$$\Rightarrow deg(f|_L) \le d.$$ ☒

(under various conditions)

Nevertheless: Thm: $\Pr_L(deg(f|_L) \le d) \le \epsilon \Rightarrow \delta(f, RM) \le 2\epsilon$. ☒

General Questions:

Low-Query Regime:

1. What is best relationship between $[n, k, d]_2$ if we want code to (ϵ, ℓ) ~~locally decodable~~ LDC with $\epsilon = \Omega(1)$ & $\ell \leq 2, 3, 4 \ldots$

   $d = \Omega(n)$; $n \geq k^{1+\frac{1}{\ell}}$; $n \lessapprox \exp(\exp(\sqrt{\log k}))$

2. What is best $[\quad]_q$ if we want code to be (ϵ, ℓ) - LTC with $\epsilon = \Omega(1)$ & $\ell \leq 2, 3, 4 \ldots$

   $n = O(k \operatorname{polylog} k)$ with $\ell = 3$.

High-rate Regime

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

LDC: $\ell \lessapprox 2^{\sqrt{\log n}} \ldots$

LTC: $\ell \leq (\log n)^{\log \log n}$

[Next two lectures : Ideas]

Rest of Today : 10^9 \$ Aside

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

In Cloud Storage: two kinds of flaws:

(A) — 1 Server / Memory bank goes down periodically;

(B) — Several Servers ~~may~~ go do down rarely;

want to recover from both; However

if (A) : then want very quick recovery ("local")

if (B) : ~~slow~~ slow recovery is O'K.

error model = erasure.

(ℓ, d)- LRC code : (1) distance $\geq d$ [corrects $d-1$ erasures]

recoverable?
reconstructible?
~~repairable~~

& (2) ~~$\forall$ message symbol~~
Code $[n] = [k] \cup [n-k]$
message — $[k]$; parity — $[n-k]$

WEAK (i) $\forall$ message $i \in [k]$ $\forall$ codeword $C_1 \ldots C_n$

C_i can be recovered from C_S for some $|S| \leq \ell$, $i \notin S$

STRONG (ii) $\forall\ i \in [n]$

C_i can be recovered from C_S for some $|S| \leq \ell$, $i \notin S$

<u>Question</u> what is the relationship between n, k, ℓ, d $(q \to \infty)$

~~[GHSY]~~:

[Trivial]: weak recovery possible with

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

Given $m = \underbrace{m_1 \ldots m_\ell}_{\oplus}, \underbrace{m_{\ell+1} \ldots m_{2\ell}}_{\oplus} \cdots \underbrace{\quad m_k}_{\oplus}$

$$\text{Encoding} = m, \left\{ \bigoplus_{i=1}^{\ell} m_{\ell(j-1)+i} \right\}_{j=1}^{k/\ell}, \underbrace{RS^{\oplus}(m)}$$

Parity check bits in systematic RS code of distance d.

Locality here

distance here

[GHSY]: $n \geq k + \frac{k}{\ell} + d - O(1)$.

<u>Proof</u>: - Find blocks of size $(\ell+1)$ of rank $\leq \ell$.

- Union them to get $(k-1)\left(\frac{\ell+1}{\ell}\right)$ coordinates of rank $\leq (k-1)$

- Apply PHP to conclude

$$d \leq n - \frac{(k-1)(\ell+1)}{\ell} \quad \text{or} \quad n \geq (k-1) + \left(\frac{k-1}{\ell}\right) + d$$

[Tamo-Burg] $\exists$ Strong (ℓ, d)-codes with $n = k + \frac{k}{\ell} + d \pm O(1)$.

<u>Construction</u>: $\cancel{q = (\ell+1) \cdot q'}$ $\quad q = \underbrace{(\ell+1)}_{r} . S + 1$

$$\boxed{\text{Such } \mathbb{F}_q \text{ has } \omega \text{ s.t. } 1, \omega, \omega^2, \ldots \omega^{r-1} \text{ distinct } \ \& \ \omega^r = 1.}$$

~~Code =~~ $\cancel{\{ f(\alpha) \mid \alpha \in \mathbb{F}_q^* \mid f(x) = \sum \alpha_i x^i}$

But $\cancel{\alpha_{r-1}, \alpha_{2r}, \alpha_{3r} \cdots} = 0 \}$

Code = $\cancel{\{ (f(\alpha)) \ \alpha \in \mathbb{F}_q^*,}$

$$C_k \quad \{ \langle f(\alpha) \rangle_{\alpha \in \mathbb{F}_q^*} \mid f(x) = \sum_{i \leq k} \alpha_i x^i$$

$$\alpha_i = 0 \text{ if } i = -1 \bmod r$$

$$\alpha_{r-1}, \alpha_{2r-1}, \alpha_{3r-1} \cdots = 0 \}$$

$$\dim(C_k) = k \left(\frac{r-1}{r} \right)$$

$$\text{dist}(C_k) \geq n - k$$

Locality ?

let $S_a = \{ a, a \cdot \omega, a\omega^2, \ldots a\omega^\ell \}$

$$f|_{S_a} = f(x) \{ \bmod \prod_{b \in S_a} (x - b) \} = f(x) \bmod (x^r - a^r)$$

But $f(x) \bmod (x^r - a^r)$ has degree $\leq r-2$!

[One smaller than it should be!]

so. $f(a)$ determined by $f\big|_{S_a - \{a\}}$!

$\boxtimes$

Next Lectures

1. ~~KS~~ High Rate LDCs
2. Some analysis of LTCs.