Today's lecture

- self - correcting for linear fctns.
- testing linearity

Linear functions:

$f: G \to H$

G, H finite groups (closure, associative, identity, inverses) with operations $+_G, +_H$ respectively

f is "linear" (homomorphism) if

$\forall\, x, y \in G \quad f(x) +_H f(y) = f(x +_G y)$

examples of finite groups: $G = \mathbb{Z}_m$ with operation "+ mod m"

$= \mathbb{Z}_m^k$ with coordinatwise "+ mod m"

examples of homomorphisms:

$f(x) = x$

$f(x) = 0$

$f(x) = ax \bmod q$

$f_{\bar{a}}(x) = \sum a_i x_i \bmod 2$

$= (x_1 \cdots x_n) \cdot \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$

def. f is "ε-linear" if $\exists$ linear g s.t.

f & g $\underbrace{\text{agree on } \geq 1-\varepsilon \text{ fraction of inputs}}_{\Pr_{x \in G}[f(x) = g(x)] \geq 1-\varepsilon}$

(else, f is "ε-far" from linear)

A useful observation:

$$\forall\, a, y \in G \qquad \Pr_x [y = a + x] = \frac{1}{|G|}$$

since only $x = y - a$ satisfies the equation

$\Rightarrow$ if pick $x \in_R G$

then $a + x$ is distributed uniformly in G

i.e. $a + x \in_R G$

example: if $G = \mathbb{Z}_2^n$ with operation $(a_1 \cdots a_n) + (b_1 \cdots b_n)$
$= (a_1 \oplus b_1, a_2 \oplus b_2, \ldots, a_n \oplus b_n)$

then

$(0110) + (b_1 b_2 b_3 b_4) = (0 \oplus b_1, 1 \oplus b_2, 1 \oplus b_3, 0 \oplus b_4)$
is distributed uniformly if b_i's are

why? each coord unif
b_i's indep $\Rightarrow$ $a_i \oplus b_i$'s indep

Why do we want it?

Self-correcting (i.e. random self-reducibility)

Given f s.t. $\exists$ linear g s.t. $\Pr_x[f(x)=g(x)] \geq 7/8$.

To compute $g(x)$: (using calls to f not g)

For $i = 1 \ldots c \log \frac{1}{\beta}$

pick $y \in_R G$

$\text{answer}_i \leftarrow f(y) + f(x-y)$

↑ unif dist by observation

Output most common value for answer_i

Claim $\Pr[\text{output} = g(x)] \geq 1-\beta$

Pf

$\Pr[f(y) \neq g(y)] \leq 1/8$

$\Pr[f(x-y) \neq g(x-y)] \leq 1/8$

$\therefore \Pr[\underbrace{f(y) + f(x-y)}_{\text{answer}_i} \neq \underbrace{g(y) + g(x-y)}_{=g(x)}] \leq 1/4$

rest is Chernoff.

Linearity Testing

Goal Given f

- if f is linear, pass
- if f is ε-far from linear, fail with prob $\geq 2/3$

Proposed Test

how big should s be?

do s times:

  Pick $x, y \in_u G$

  if $f(x) + f(y) \neq f(x+y)$ output "FAIL" + halt

Output "PASS"

Behavior of test

if f linear, passes with prob 1 ✓

if f ε-far from linear?

  will prove contrapositive:

   if f likely to pass $\Rightarrow$ f is ε-linear

   (equivalent to "if f is ε-far then f is likely to fail")

Plan:

if f is close to linear,

then function g you get from self-correcting f

namely $g(x) = \text{majority}_y [\underbrace{f(x+y) - f(y)}_{\text{y's vote for } x}]$

will be (1) linear

(2) close to f.

if f is not close to linear, then no guarantees

would like to show that if test fails rarely,

then you do get guarantees!

for example:

(1) most x satisfy $f(x) = \text{majority}_y [f(x+y) - f(y)]$ overwhelmingly

(2) if x, y satisfies

then maybe $x+y$ also satisfies ?

& maybe we can say something about

$g(x+y) = g(x) + g(y)$?

<u>Thm</u> Suppose $\delta \equiv \Pr_{x,y}[f(x)+f(y) \neq f(x+y)] < \frac{1}{16}$. Then f is $\underset{\varepsilon}{2\delta}$-close to linear.

$\Rightarrow$ s needs to be big enough to verify for $\delta < 1/16$, so need $s \gg 16$ & $s = \Omega(\frac{1}{\delta}) = \Omega(\frac{1}{\varepsilon})$

<u>Proof.</u>

g is self-correction of f on x $\Rightarrow$

<u>def</u> $g(x) \equiv \text{plurality}_y [\underbrace{f(x+y)-f(y)}_{\text{y's vote for } f(x)}]$ $\leftarrow$ break ties arbitrarily

<u>def</u> x is ρ-good if $\Pr_y[\underbrace{g(x) = f(x+y)-f(y)}] \geq 1-\rho$

else ρ-bad

i.e. $> 1-\rho > \frac{1}{2}$ fraction of y's agree on their vote

x is ρ-good for $\rho < \frac{1}{2}$ $\Rightarrow$ $g(x)$ defined via majority element

First: Show g & f agree usually

<u>Claim 1</u> $\rho < \frac{1}{2}$

$\Pr_x[x \text{ is } \rho\text{-good} \;\&\; g(x) = f(x)] > 1 - \delta/\rho$ $\Rightarrow$ fraction of x for which f & g agree is $> 1-2\delta > \frac{7}{8}$

<u>Pf of claim 1</u>

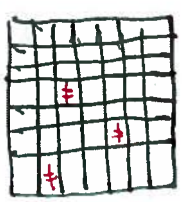
$\alpha_x = \Pr_y[f(x) \neq f(x+y) - f(y)]$

if $\alpha_x \leq \rho < \frac{1}{2}$ then x is ρ-good & $g(x) = f(x)$

Picture of proof

all y's

all x's



Matrix: fraction of "$\neq$" entries $= \delta$

$E[\text{fraction } \text{"}\neq\text{" entries in row}] = \delta$

Fraction rows with $> c\cdot\delta$ fraction entries has to be $< \frac{1}{c}$ by Markov's $\neq$ $\rightarrow$

Use Markov's $\neq$:

$$E_x[\alpha_x] = \frac{1}{|G|}\sum_{x \in G} \Pr_y[f(x) \neq f(x+y) - f(y)]$$

$$= \Pr_{x,y}[f(x) \neq f(x+y) - f(y)]$$

$$= \delta$$

so $\Pr[\alpha_x > \rho] \leq \frac{\delta}{\rho}$

$\| $

$(\frac{\rho}{\delta})\delta$

Second: Show g "is a homomorphism" (at least where it is defined)

<u>Claim 2</u> $\rho < 1/4$

if x, y both ρ-good then (at least $3/4$ x's are $1/4$-good)

(1) $x+y$ is 2ρ-good

(2) $g(x+y) = g(x) + g(y)$

<u>Pf of Claim 2</u>

let $h(x+y) = g(x) + g(y)$

$\Pr_z[g(y) \neq f(y+z) - f(z)] < \rho$ since y is ρ good

$\Pr_z[g(x) \neq f(x+(y+z)) - f(y+z)] < \rho$ since x is ρ-good & $y+z \in_u G$

So $\Pr_z[h(x+y) = g(x) + g(y)$ by def

$= f(x+(y+z)) - \cancel{f(y+z)} + \cancel{f(y+z)} - f(z)] \geq 1-2\rho > 1/2$

Union bnd using

$\Downarrow$

$g(x+y) = h(x+y)$ by def of g (since $f(x+y+z) - f(z)$ is <u>same</u> for $\geq \frac{1}{2}$ of z's!)

$= g(x) + g(y)$ " " " h

& $x+y$ is 2ρ-good ∎

Third: show g is defined for all x

Claim 3 $\delta < 1/16$

$\forall x,\quad x$ is 4δ-good ($\frac{1}{4}$-good) & $g(x)$ defined via majority elt.

Pf.

if $\exists y$ st. y & $x-y$ both 2δ-good

claim 2 $\Rightarrow$ x is 4δ-good

& $g(x) = g(y) + g(x-y)$

but $\Pr_y[\underbrace{y \text{ \& } (x-y)}_{\text{both uniform}}$ both 2δ-good$] > \underbrace{1 - \underbrace{\left(\frac{\delta}{2\delta}\right)}_{\text{claim 1}} \cdot 2}_{\text{union bnd}} = 0$

$\Rightarrow \exists y$ st. y & $(x-y)$ both 2δ-good ∎

Claim 3 $\Rightarrow$ g defined $\forall x$ as majority elt.

By claim 2, $\forall x, y \quad g(x) + g(y) = g(x+y)$

By claim 1, f & g agree $\geq 1 - 2\delta$ fraction of G ∎

Improved theorem:

only need $\delta < 2/9$

(this means $O(9/2)$ many tests give < const prob of failure, instead of $O(16)$ – is this a big deal? actually it can be ...)

$2/9$ is tight: there are fctns that are far from linear but pass test with prob $7/9$

Coppersmith's example:

$$f(x) = \begin{cases} 1 & \text{if } x = 1 \bmod 3 \\ 0 & \phantom{\text{if }} 0 \\ -1 & \phantom{\text{if }} 2 \end{cases}$$

integers over $\mathbb{Z}$

$f(x) + f(y) = 2$
$f(x+y) = -1$

f fails when $x = y = 1 \bmod 3$
$x = y = 2 \bmod 3$ $\Big\}$ Prob $= 2/9$ ← not bad!

else passes

Closest linear fctn is $f(x) \equiv 0$ ← $\Pr[f(x) = g(x)] = 1/3$ very far!!
$\varepsilon = 2/3$

$\delta = 2/9$ is a "threshold"