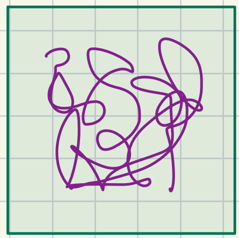
Testing + Correcting Linear Functions

Program Correctness:

Suppose you have a program for fctn f



complicated spaghetti code – highly optimized

why should you trust it?

"Self-Correcting" – take a program that is correct on most inputs & transform into program correct on all inputs.

"self-testing" – convince yourself that program is correct on most inputs

You can do this for certain classes of fctns f!

Linear Functions

closure, associative, identity, inverses

$f: G \to H$ $\quad$ G, H finite groups with operators $+_G, +_H$ respectively

def. f is "linear" (homomorphism) if

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

examples of finite groups:

$G = \mathbb{Z}_m$ $\quad$ (#'s mod m) with operation "+ mod m"

$G = \mathbb{Z}_m^k$ $\quad$ k-vector with coordinatewise "+ mod m"

examples of homomorphisms: multiplication, division...

$f: G \to G$ $\quad$ $f(x) = x$ $\quad$ $\underbrace{f(x) = a \cdot x \bmod q}_{\text{for } G = \mathbb{Z}_q}$ $\quad$ multiplication

$f(x) = 0$

$$f_{\bar{a}}: \mathbb{Z}_2^k \to \mathbb{Z}_2 \qquad f_{\bar{a}}(x) = \sum a_i x_i \bmod 2 = (x_1 \cdots x_k) \cdot \begin{pmatrix} a_1 \\ \vdots \\ a_k \end{pmatrix}$$

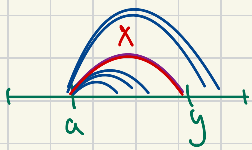
<u>def</u>. f is "ε-linear" if $\exists$ linear fctn g

s.t. $\underbrace{f \text{ \& } g \text{ agree on } \geq 1-\varepsilon \text{ fraction of inputs}}$

write as: $\Pr_{x \in G}[f(x) = g(x)] \geq 1-\varepsilon$

else f is "ε-far from linear"

A useful observation:



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

since only $x = y - a$ satisfies equation

$\Rightarrow$ if pick $x \in_R G$ then $a+x$ distributed uniformly in G

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

example: if $G = \mathbb{Z}_2^n$ with operation

$$(a_1 \cdots a_n) + (b_1 \cdots b_n) = (a_1 \oplus b_1, \ldots, a_n \oplus b_n)$$

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 care?

Self-Correcting (ie. random self-reducibility)

Given: f s.t. $\underbrace{\exists \text{ linear } g \text{ s.t. } \Pr_x[f(x) = g(x)] \geq 7/8}_{f \text{ is } \frac{1}{8}\text{-linear}}$

Compute: $g(x)$ given oracle calls to f

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

pick $y \in_R G$

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

uniform since y is + by observation

Output most common value for answer_i

hope $g(x) = g(y) + g(x-y)$ since g linear

if $f(y) = g(y)$ & $f(x-y) = g(x-y)$ then

will output $g(x)$

runtime: $O(\log \frac{1}{\beta})$ calls to f

note: learning f takes many more calls to f

(on $\mathbb{Z}_2^n$ need n calls)

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

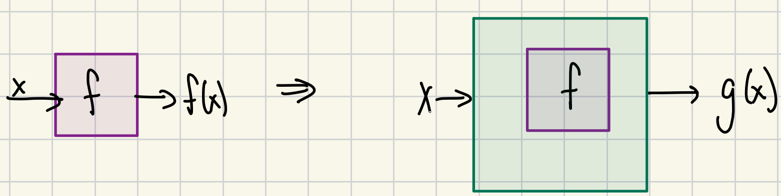
Pf

$\Pr[f(y) \neq g(y)] \leq 1/8$ since y uniform & f ε-close to g

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

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

Chernoff $\Rightarrow$ majority answer is $g(x)$ with prob $\geq 1-\beta$ ∎



Linearity Testing

how do we know that f is ε-linear?

Goal: given f

- if f linear, pass
- if f $\underbrace{\varepsilon\text{-far}}$ from linear, fail with prob $\geq 2/3$

  need to change f on $\geq \varepsilon$ fraction of domain

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, always passes

if f ε-far, show contrapositive:

f passes whp $\Rightarrow$ f close to linear

Plan

- if f ε-close to linear then fctn g you get from self-correcting f, namely

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

will be (1) linear

(2) close to f

- if f <u>not</u> close to linear, then no guarantees on $g(x)$

<u>but</u> if test rarely fails, then you do get guarantees

e.g. • most x satisfy $f(x) = \text{maj}_y [f(x+y) - f(y)]$

• for such x, y maybe $x+y$ also satisfies it?

Let $\delta \equiv \Pr_{x,y}\left[f(x)+f(y) \neq f(x+y)\right]$

fraction of pairs x,y which fail test written as probability

<u>Thm</u> Suppose $\delta < 1/16$. Then f is $\underbrace{2\delta}_{\varepsilon}$-close to linear.

<u>Note</u> to ensure $\delta < \delta_0$, need only $O(1/\delta_0)$ tests.

$\Rightarrow$ to ensure f is ε-linear, need only $O(1/\varepsilon)$ calls to f.

<u>Proof of Thm</u>

Let g be self-correction of f

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

Say g is "winner" if $g(x)$ agrees with majority vote } how often does it happen?

For $\rho < 1/2$:

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

else "ρ- bad"

$>\underbrace{1-\rho}_{>1/2}$ fraction of y's agree on vote

$\rho < 1/2 \Rightarrow$ "winner" $\left\{ \begin{array}{l} g(x) \text{ defined via} \\ \text{majority element} \end{array} \right.$

1^{st}: g + f usually agree + lots of winners!

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

$\Pr_x[x \text{ is } \rho\text{-good} + g(x) = f(x)] > 1 - \delta/\rho$

<u>Corr</u> fraction of x for which f + g agree is $\underset{\substack{\uparrow \\ \rho < 1/2}}{>} 1 - 2\delta > 7/8$

Pf of Claim 1

let $\alpha_x \equiv \Pr_y[f(x) \neq f(x+y) - f(y)]$ ← fraction of $\neq$ in a row

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

argument via picture

Consider matrix:

$$\begin{array}{cc} & \text{all } y\text{'s} \\ \text{all } x\text{'s} & \begin{array}{|c|c|c|c|} \hline = & = & \neq & = \\ \hline \neq & \neq & = & = \\ \hline = & \neq & = & = \\ \hline = & = & = & = \\ \hline \end{array} \end{array}$$

$\neq$ if $f(x) + f(y) \neq f(x+y)$

$=$ o.w.

a row is ρ-good if $\leq \rho$ fraction of $\neq$'s

fraction of $\neq$ in matrix $= \delta$

$E_{\text{row}}[\text{fraction of } \neq \text{ in row}] = \delta$

fraction of rows w/ $> c \cdot \delta$ $\neq$'s is $\leq \frac{1}{c} \cdot \delta$

(Markov's inequality)

more formally:

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

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

$$= \delta$$

So $\Pr[\alpha_x > \underbrace{\rho}_{= (\rho/\delta) \cdot \delta}] \overset{\text{Markov's}}{\leq} \delta/\rho$ ∎

Show g is a "homomorphism" (at least where "well defined")

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

if x, y both ρ-good then

(1) $x+y$ is 2ρ-good

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

<u>Pf of claim 2</u>

let $h(x+y) \equiv g(x) + g(y)$

two "bad" events unlikely:

$$\Pr_z[g(y) \neq f(y+z) - f(z)] < \rho \quad \text{since } y \text{ is } \rho\text{-good}$$

$$\Pr_z[g(x) \neq f(x+(y+z)) - f(y+z)] < \rho$$

since x is ρ-good & $(y+z)$ is uniform

so $\Pr_z[h(x+y) \equiv g(x) + g(y)$

$= f(x+y+z) - f(y+z) + f(y+z) - f(z)$

$= f(x+y+z) - f(z)] \underset{\uparrow \text{union bnd}}{>} 1 - 2\rho \underset{\uparrow \rho < 1/4}{\geq} 1/2$

So there is some value, namely $h(x+y)$,

which is equal to $f(x+y+z) - f(z)$

for a majority of z's

$\Rightarrow h(x+y) = g(x+y)$ (def of g)

but since $h(x+y) = g(x) + g(y)$ (def of h)

we have $g(x+y) = g(x) + g(y)$

also, $(x+y)$ is 2ρ-good. ∎

Show g "well-defined" for all x.

<u>Claim 3</u> $\delta < 1/16$

$\forall x$, x is 4δ-good & $g(x)$ defined

via majority element.

Claim 3 $\Rightarrow$ $\forall x$, g "well-defined" ← winner

Claim 2 $\Rightarrow$ $\forall x, y$ $g(x+y) = g(x) + g(y)$

Claim 1 $\Rightarrow$ g & f agree on $\geq 1-2\delta$ fraction of G.

Pf of Claim 3

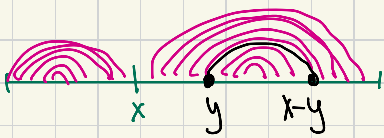
if $\exists y$ s.t. both y & $(x-y)$ are 2δ-good

then claim 2 $\Longrightarrow$ $x = y + (x-y)$ is 4δ-good

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

To show such y exists:

pair off all y with $(x-y)$



$$\Pr_y[y + (x-y) \text{ both } 2\delta\text{-good}] > 1 - 2 \cdot \frac{\delta}{2\delta} > 0$$

claim 1

Since prob > 0, $\exists$ pair $y, (x-y)$ both 2δ-good $\blacksquare$

Finished proof of theorem!!!

Comments

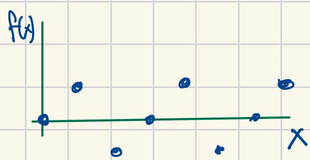
only need $\delta < 2/9$

($\Rightarrow$ $O(9/2)$ tests instead of $O(1/\delta)$

why do we care?)

actually turns out to be a "threshold" $\rightarrow$ $2/9$ is tight:

$\exists$ fctns far from linear but pass test with prob $7/9$



Coppersmith's example:

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

f fails when $\left.\begin{array}{l} x=y=1 \mod 3 \\ x=y=2 \mod 3 \end{array}\right\}$ prob $2/9$

else passes

Closest linear fctn is $g(x) \equiv 0$ which disagrees with f on $\geq 2/3$ of pts.

more comments:

- above proof requires underlying group to be Abelian $(x+y = y+x)$ but can prove for non-Abelian
- Better constants for $\mathbb{Z}_2^n$

why do we care?

PCP constructions