6.889 Sublinear Time Algorithms May 1, 2019

Lecture 22

Lecturer: Ronitt Rubinfeld *Scribe: Bertie Ancona*

1 Introduction

Today we will go over linear functions,how to self-correct them and how to test them.

Definition 1 *A function $f : G \to H$, where G and H are finite groups having operations $+_G$ and $+_H$, is* linear *(homomorphic) if $f(x) +_H f(y) = f(x +_G y)$ for all $x, y \in G$.*

Examples of finite groups:

- Z_m with addition mod m
- Z_m^k with coordinate-wise addition mod m

Examples of linear functions:

- $f(x) = 0$
- $f(x) = x$
- $f(x) = ax \mod m$
- $f_{\bar{a}}(\bar{x}) = \sum_i a_i x_i \mod m$

Definition 2 *A function f is ϵ-*linear *if there is some linear function g such that f and g agree on an $(1-\epsilon)$ fraction of inputs. Otherwise, f is ϵ-*far *from linear.*

This is equivalent to having $\Pr_{x\in G}[f(x) = g(x)] \geq 1 - \epsilon$.

A Useful Observation For all $a, y \in G$, $\Pr_{x\in G}[y = a+x] = \frac{1}{|G|}$, because only a single value $x = y - a$ satisfies this. Thus, if $x \in_R G$ (x chosen from G uniformly at random), then $a + x \in_R G$ for all $a \in G$.

2 Self-Correction (or, Random Self-Reducibility)

Given a function f such that f is $\frac{1}{8}$-linear, let g be a linear function $\frac{1}{8}$-close to f. To compute $g(x)$:

Algorithm 1 Self-Correcting

```
for i in 1, ..., c log(1/β) do
    Pick y ∈_R G
    answer_i ← f(y) + f(x − y)
end for
Output most common value over all answer_i
```

Claim 3 *After running Algorithm 1,* $\Pr[\mathit{Output} = g(x)] \geq 1 - \beta$

Proof $\Pr[f(y) \neq g(y)] \leq \frac{1}{8}$ (by definition)

$\Pr[f(x-y) \neq g(x-y)] \leq \frac{1}{8}$ (by our Useful Observation)

$\Rightarrow \Pr[f(y) + f(x-y) \neq g(y) + g(x-y)] = \Pr[\mathit{answer}_i \neq g(x)] \leq \frac{1}{4}$ (by linearity and union bound)

Now we may use Chernoff to show that most common value of answer_i will be $g(x)$ with probability $1 - \beta$ after $c \log \frac{1}{\beta}$ iterations. ■

3 Testing

The Goal: Given f, if f is linear then PASS with probability 1. If f is ϵ-far from linear, FAIL with probability at least 2/3.

Algorithm 2 Linearity Testing

```
for s times do
    Pick x, y ∈_R G
    if f(x) + f(y) ≠ f(x + y) then
        Output FAIL and halt
    end if
end for
Output PASS and halt
```

If f is linear, Algorithm 2 clearly passes with probability 1. We will prove the contrapositive for *eps*-far f: if f is likely to pass, then f is ϵ-linear.

Theorem 4 *Say* $\delta = \Pr_{x,y}[f(x) + f(y) \neq f(x+y)] < \frac{1}{16}$. *Then* f *is* 2δ*-linear.*

This would mean that setting $s = \Omega(1/\delta) = \Omega(16)$ is enough for such f to be likely to pass Algorithm 2.
Proof

Definition 5 *Let* $g(x) = plurality_y\{f(x+y) - f(y)\}$, *breaking ties arbitrarily.*

In other words, $g(x)$ is the self-correction of f on x.

Definition 6 x *is* ρ*-good if* $\Pr_y[g(x) = f(x+y) - f(y)] \geq 1 - \rho$ *(i.e., a* $(1-\rho)$ *fraction of* y*'s agree on their vote for* $f(x)$*), and* x *is* ρ*-bad otherwise.*

This means that if x is $\frac{1}{2}$-good, then $g(x)$ is defined on the majority element.

We prove Theorem 4 in three claims. With Claim 9, we show that g is defined for all x as the majority element. With Claim 8, we show that g is "linear". Finally, with Claim 7 we show that f and g agree on at least a $1 - 2\delta$ fraction of inputs, i.e. that they are 2δ-close, implying that f is 2δ-linear. We now prove the claims.

Claim 7 *If* $\rho < \frac{1}{2}$, $\Pr_x[x$ *is* ρ*-good and* $g(x) = f(x)] > 1 - \frac{\delta}{\rho}$

The claim implies that the fraction of x for which f and g both agree is greater than $1-\delta/\rho > 1-2\delta > 7/8$.
Proof

Let $\alpha_x = \Pr_y[f(x) \neq f(x+y) - f(y)]$.
If $\alpha_x \leq \rho < 1/2$, then x is ρ-good and $g(x) = f(x)$ (and we have our claim).
$\mathrm{E}_x[\alpha_x] = \frac{1}{|G|}\sum_{x \in G} \Pr_y[f(x) \neq f(x+y) - f(y)]$
$= \Pr_{x,y}[f(x) \neq f(x+y) - f(y)]$
$= \delta$. Now by Markov:
$\Pr[\alpha_x > \rho] \leq \frac{\delta}{\rho} \Rightarrow \Pr[\alpha_x \leq \rho] \geq 1 - \frac{\delta}{\rho}$. ■

Claim 8 *If* $\rho < \frac{1}{4}$ *and* x *and* y *are both* ρ*-good, then (1)* $x+y$ *is* 2ρ*-good, and (2)* $g(x+y) = g(x) + g(y)$.

Proof Let $h(x, y) = g(x) + g(y)$.

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

$\Pr_z[g(x) \neq f(x + (y+z)) - f(y+z)] < \rho$ (because x is ρ-good and $(y+z) \in_R G$). We have that $h(x, y) = g(x) + g(y)$, therefore

$\Pr_z[h(x, y) = f(x + (y+z)) - f(y+z) + f(y+z) - f(z) \equiv f((x+y)+z)) - f(z)] > 1 - 2\rho > \frac{1}{2}$ (by union bound of the above).

This means that $g(x+y) = h(x, y)$, because $f((x+y)+z)) - f(z)$ is more than half of the votes and thus wins plurality for $g(x+y)$, by definition of g.

Also, $h(x, y) = g(x) + g(y)$ by definition of h, so $g(x+y) = g(x) + g(y)$. We also have that $(x+y)$ is 2ρ-good by the last probability statement. ■

Claim 9 *If $\delta < \frac{1}{16}$, then for all x, x is 4δ-good and $g(x)$ is defined as the majority element.*

Proof If there is a y such that y and $x+y$ are both 2δ-good, then by claim 8, x is 4δ-good and $g(x) = g(y) + g(x-y)$.

We prove that such a y must exist.

$\Pr_y[y$ and $x+y$ are both 2δ-good$] > 1 - 2(\frac{\delta}{2\delta}) = 0$, by claim 7 and union bound. Thus, such a y must exist and the claim holds. ■

■

3.1 δ Tightness

It is in fact possible to show this for $\delta < \frac{2}{9}$, rather than $\delta < \frac{1}{16}$. We show that we cannot do better than $\frac{2}{9}$ with an example of a function that is $\frac{2}{3}$-far from linear but passes our test with probability $\frac{7}{9}$.

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

The closest linear function is $g(x) = 0$, which is $\epsilon = \frac{2}{3}$-far from f. However, our test only fails in two of nine cases:

- When $x = y = 1 \mod 3$, $f(x) + f(y) = 2 \mod 3$ and $f(x+y) = -1 \mod 3$
- When $x = y = 2 \mod 3$, $f(x) + f(y) = -2 \mod 3$ and $f(x+y) = 1 \mod 3$