Today:

Learning large Fourier coefficients with <u>queries</u>

[Goldreich Levin]

[Kushilevitz - Mansour]

$$x \to \boxed{f} \to f(x)$$

Given $f, \theta$

1) Output all $S$ s.t. $|\hat{f}(S)| \geq \theta$ ← all close parity fctns

2) All outputted $S$ satisfy $|\hat{f}(S)| \geq \frac{\theta}{2}$ ← no junk

(Probably) can't do it with random examples

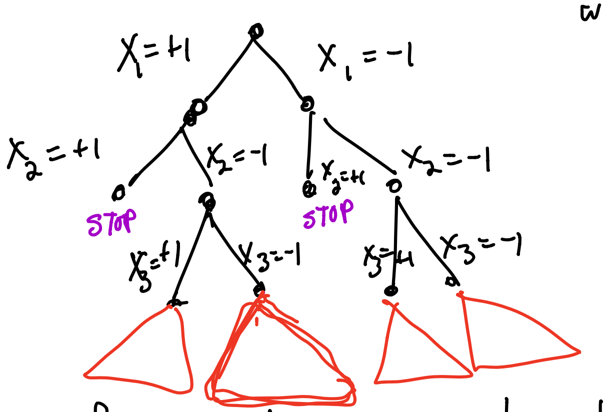
do queries help?

Last time: (warmup) only one big $\hat{f}(S)$

Algorithm: find $S$ bit-by-bit

## General Case Main Idea:

exhaustive search with good pruning



idea: find way to use sampling to estimate "total energy" in subtree

only go down high energy paths

## How do we prune?

Define useful quantity:

Fix $0 \leq k \leq n$ current level of search

$S_1 \subseteq [k]$ current "node" of search

$S_1 \sim$ rt turns

$$f_{k,S_1} : \{\pm 1\}^{n-k} \to \mathbb{R}$$

s.t. $f_{k,S_1}(z) = \underbrace{\sum_{T_2 \subseteq \{k+1,..n\}} \hat{f}(S_1 \cup T_2)\chi_{T_2}(z)}_{}$

will call with $y = x_{k+1} .. x_n$

index 1 ~ prefix
index 2 ~ suffix

Four coeffs $\hat{f}(S)$
s.t. $S$ agrees
with $S_1$ on $[k]$

Sanity checks:

1) $k=0$ $\quad f_{0,\emptyset} = \sum_{T_2 \subseteq [n]} \hat{f}(T_2)\chi_{T_2}(x) = f(x)$

2) $k=n$ $\quad f_{n,S_1}(x) = \hat{f}(S_1)$

Partition Four coeff into $2^k$ subsets

Plan Go down paths with $\underbrace{E[f^2_{k,S}(x)] \geq \theta^2}_{\text{why?}}$

1. Can compute?
2. Does it bring us to right leaves?
   - do we get all heavy leaves?
   - do we get too much junk? (light leaves)
3. How many paths do we take?
   lots of deadends?
   runtime good?

## Not too many paths (answer to 3)

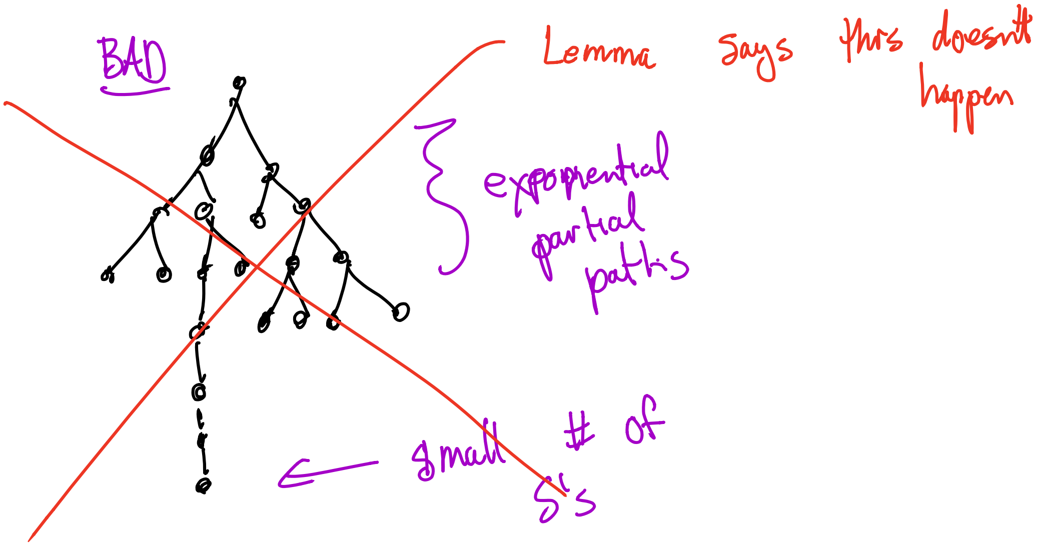
**Lemma** "not too many" — not just at end but also at any stage (level of tree)

$f$ Boolean

(1) $\leq \frac{1}{\theta^2}$ $S$'s satisfy $|\hat{f}(S)| \geq \theta$

(2) $\forall\ 0 \leq k \leq n$, $\leq \frac{1}{\theta^2}$ fctns $f_{k,S_k}$ have

$$E_x[f^2_{k,S_1}(x)] \geq \theta^2$$



Pf. (1) Boolean Parseval's: $1 = E_x[f^2(x)] = \sum_S \hat{f}(S)^2$

So if $> \frac{1}{\theta^2}$ $S$'s st. $|\hat{f}(S)| \geq \theta$

$$\Rightarrow \sum_S \hat{f}(S)^2 > \frac{1}{\theta^2} \cdot \theta^2 > 1 \rightarrow\leftarrow$$

(2) For given k:

Claim: $\forall\ k,\ S_1 \subseteq [k]$

$$E_x[f_{k,S_1}(x)^2] = \sum_{T_2 \subseteq \{k+1,\dots,n\}} \hat{f}(S_1 \cup T_2)^2$$

Pf

$$E_x[f_{k,S_1}(x)^2] = E_x\Big[\Big(\sum_{T_2} \hat{f}(S_1 \cup T_2)\chi_{T_2}(x)\Big)^2\Big] \quad \text{def of } f_{k,S_1}$$

$$= \sum_{T_2,T_2'} \hat{f}(S_1 \cup T_2)\hat{f}(S_1 \cup T_2')\, E_x\big[\underbrace{\chi_{T_2}(x)\chi_{T_2'}(x)}_{\substack{=1 \text{ if } T_2=T_2' \\ =0 \text{ o.w}}}\big]$$

$$= \sum_{T_2} \hat{f}(S_1 \cup T_2)^2$$

∎

Using claim:

$$1 = \sum_S \hat{f}(S)^2 = \sum_{S_1 \subseteq [k]} \sum_{T_2 \subseteq \{k+1..n\}} \hat{f}(S_1 \cup T_2)^2$$

$$= \sum_{S_1} E_x[f_{k,S_1}(x)]^2 \quad \leftarrow \text{claim}$$

as before $\le \frac{1}{\theta^2}$ $S_1$'s can have $E_x[f^2_{k,S_1}(x)] > \theta^2$

∎

$\Rightarrow \leq \frac{1}{\theta^2}$ "live" $S$'s at each level of tree

$\Rightarrow \leq \frac{n}{\theta^2}$ paths are traversed in tree

$\leq 2^n$

Answer to 3

Does it bring us to good leaves? (Answer to 2)

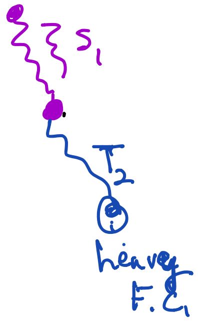
Fact "not missing out" $\Rightarrow$ find all big Four coeffs

for any $S_1$ if $\exists\, T_2$ s.t.

heavy F.C's $\{ \; |\hat{f}(S_1 \cup T_2)| > \theta$

$$\text{then } \forall k \; E_x[f^2_{k,S_1}(x)] = \sum_{T_2} \hat{f}(S_1 \cup T_2)^2$$

$$\geq \theta^2$$



$\Rightarrow E_x[f^2_{k,S_1}(x)]$ is a good measure to use to find heavy F.C's

Great we find good leaves.

But do we output bad leaves?

1) note we don't reach too many leaves in total

$\Rightarrow$ can't be too many "candidates" bad leaves

2) Can test a leaf to see if good

Given candidates $S$:

is $\hat{f}(S)$ big?

previous lectures $\Big\{$ Can estimate $\hat{f}(S)$ for any specific $S$ quickly

$\Rightarrow$ Can make sure not to output junk

---

Can we estimate $f_{k,S_1}(x)$ (answer to (1))

Bad idea: estimate each $\hat{f}(S_1 \cup T_2)$ $\forall T_2$

(too many $T_2$'s)

## Better idea:

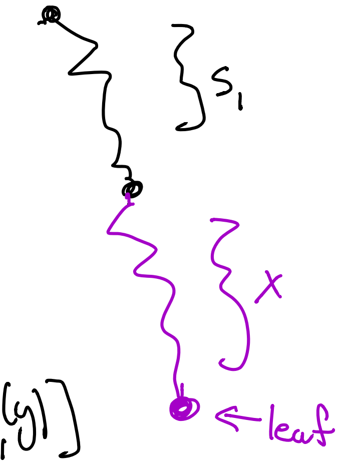
"$f_{k,S_1}(x)$ estimation" lemma:

$f : \{\pm1\}^n \to \{\pm1\}$

$0 \le k \le n$

$S_1 \subseteq [k]$

for $x \in \{\pm1\}^{n-k}$



$$\sum_{T_2 \subseteq \{k+1..n\}} \hat{f}(S_1 \cup T_2)\chi_{T_2}(x) \overset{\text{def}}{=} f_{k,S_1}(x) = \mathbb{E}_{y\in\{\pm1\}^k}[f(\underset{\uparrow}{yx})\chi_{S_1}(y)]$$

concatenation

Pf. $f(yx) = \sum_T \hat{f}(T)\chi_T(yx)$

$T = T_1 \cup T_2 \qquad T_1 \subseteq [k] \qquad T_2 \subseteq \{k+1,\dots n\}$

$\chi_T(yx) = \chi_{T_1}(y)\cdot\chi_{T_2}(x)$

$$\mathbb{E}_{y\in\{\pm1\}^k}[f(yx)\chi_{S_1}(y)]$$

$$= \mathbb{E}_y[(\sum_T \hat{f}(T)\chi_T(yx))\chi_{S_1}(y)]$$

$$= \mathbb{E}_y[\sum_{T_1}\sum_{T_2} \hat{f}(T_1 \cup T_2)\chi_{T_1}(y)\cdot\chi_{T_2}(x)\cdot\chi_{S_1}(y)]$$

$$= \sum_{T_1} \sum_{T_2} \hat{f}(T_1 \cup T_2) \cdot \chi_{T_2}(x) \underbrace{E_y[\chi_{T_1}(y) \cdot \chi_{S_1}(y)]}_{\substack{= 0 \text{ if } T_1 \neq S_1 \\ 1 \text{ if } T_1 = S_1}}$$

$$= \sum_{T_2} \hat{f}(S_1 \cup T_2) \chi_{T_2}(x)$$

$$= f_{k, S_1}(x) \qquad \blacksquare$$

Can estimate $\underbrace{E_y}_{\text{pick random } y\text{'s}} [\underbrace{f(yx)}_{\text{queries to } f} \underbrace{\chi_{S_1}(y)}_{\text{compute yourself}}]$

for several $y$'s, take average + Chernoff ✓

$O\left(\frac{1}{\gamma^2} \log \frac{1}{\delta}\right)$ queries
for add error $\gamma$
prob of bad approx $\leq \delta$

$\Rightarrow$ Answer to question 1 is "yes"

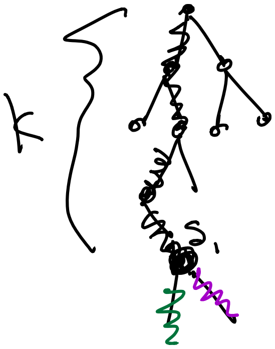
Algorithm:

current level   choices so far

~~Ideal~~ Algorithm: given $k, S_1$

leaves

If $k=n$:
test $S_1$ & output if good



Else

if $E_x[f^2_{k,S_1\cup\{k+1\}}(x)] \geq \theta^2/2$

recurse on $(k+1, S_1 \cup \{k+1\})$

(else kill this subtree)

if $E_x[f^2_{k,S_1}(x)] \geq \theta^2/2$

recurse on $(k+1, S_1)$

(else kill this subtree)

Problem only get estimate of

$$E_x[f^2_{k,S_1}(x)]$$

using $\theta^2/2$ ensures we get all heavy coeffs

using same analysis as before, still
won't have too many paths

+ still test all candidate S before
output so
no junk is output

<u>Thm</u> $f : \{\pm 1\}^n \to \{\pm 1\}$

$\forall \theta > 0$ KM-algorithm outputs set

$\mathcal{S} = \{S_1, \dots S_\ell\}$ s.t. $\ell = O(\frac{1}{\theta^2})$

s.t. with prob $\geq 1-\delta$

$\forall S_i \in \mathcal{S}$, $|\hat{f}(s)| \geq \frac{\theta}{2}$

$\forall S_i \notin \mathcal{S}$, $|\hat{f}(s) \leq \theta$

What are <u>values</u> of $\hat{f}(S)$'s ?

we can easily approximate
(previous lectures)

## Applications

- Decision trees of $\underbrace{\text{size}}_{\substack{\text{previous bnd}\\ \text{in terms}\\ \text{of depth}}} \le t$

- all fctns of small $L_1$ norm

$$L_1(f) = \sum_S |\hat{f}(S)|$$

by setting $\theta \leftarrow \frac{\varepsilon}{L_1(f)}$

in time $\text{poly}(n, L_1(f), \frac{1}{\varepsilon})$

if don't know $L_1(f)$:

guess $1, 2, 4, 8, \ldots$

run algorithm, get hypothesis,

test hypothesis &

continue if not good