Lecture 18:

Lower bound techniques

How to prove lower bounds?

easy? sublinear time algorithms see very little of input.

difficult? sublinear time algorithms are usually randomized

How to prove lower bounds?

easy? sublinear time algorithms see very little of input.

difficult? sublinear time algorithms are usually randomized

Useful lower bound tool:

Yao's Principle: Given distribution D on union of "positive" (Yes, PASS) instances + "negative" (No, FAIL) inputs, such that any deterministic algorithm of query complexity $\leq t$ is incorrect with prob $\geq 1/3$ on inputs chosen from D, then t is a lower bound on randomized query complexity.

(Proof Omitted)

ave case deterministic l.b.

$\Downarrow$

randomized worst case l.b.

Game theoretic view:

Alice selects deterministic alg A
Bob selects input x
$\Big\}$ payoff = cost of $A(x)$

A selects randomized algorithm $\Longleftrightarrow$ A picks random deterministic algorithm
(fixed once you set random bits)

Von Neuman's minimax $\Rightarrow$ when A randomized,
a $\underbrace{\text{randomized Bob}}_{\text{distribution on inputs}}$ can do just as well
as when A deterministic

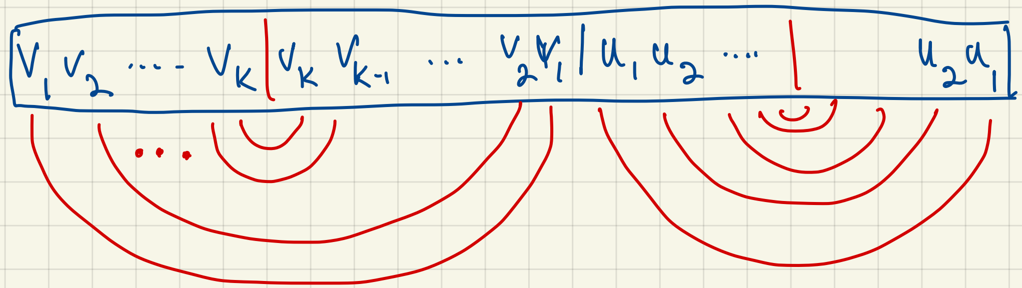
$\Rightarrow$ if want to show lower bnd, need only show distribution on inputs that is bad for every deterministic algorithm

Example application of Yao's method:

$$2PAL = \{ w \mid w \text{ is } w = v v^R u u^R \}$$

Concatenation of palindromes

e.g. $\underbrace{000111}_{v}\,\underbrace{111000}_{v^R}\,\underbrace{011100}_{u}\,\underbrace{001110}_{u^R}$



Note that testing $PAL = \{ w \mid w = v v^R \}$ is trivial:

pick random i, if $w_i \neq w_{n-i}$ fail

Thm any property tester for 2PAL needs $\sqrt{n}$ queries

e.g. if $\mathcal{A}$ satisfies $\forall x \in 2PAL$, $\Pr[\mathcal{A}(x) = PASS] \geq 2/3$

& $\forall x$ ε-far from 2PAL, $\Pr[A(x) = FAIL] \geq 2/3$

then $\mathcal{A}$ makes $\Omega(\sqrt{n})$ queries

Pf. Plan: give distribution on inputs that is hard for all algorithms using $o(\sqrt{n})$ queries.

Yao $\Rightarrow$ randomized l.b. of $\Omega(\sqrt{n})$

Distribution on "Fail" inputs:

F = random string of distance $\geq \varepsilon n$ from 2PAL

Distribution on "Pass" inputs: (wlog assume $6 | n$)

$$P = \begin{cases} 1. & \text{pick } k \in_R \left[\frac{n}{6}+1, \frac{n}{3}\right] \\ 2. & \text{pick random } v, u \text{ s.t. } |v| = k, \ |u| = \frac{n-k}{2} \\ 3. & \text{output } v v^R u u^R \end{cases}$$

note: some strings can be generated via multiple k's

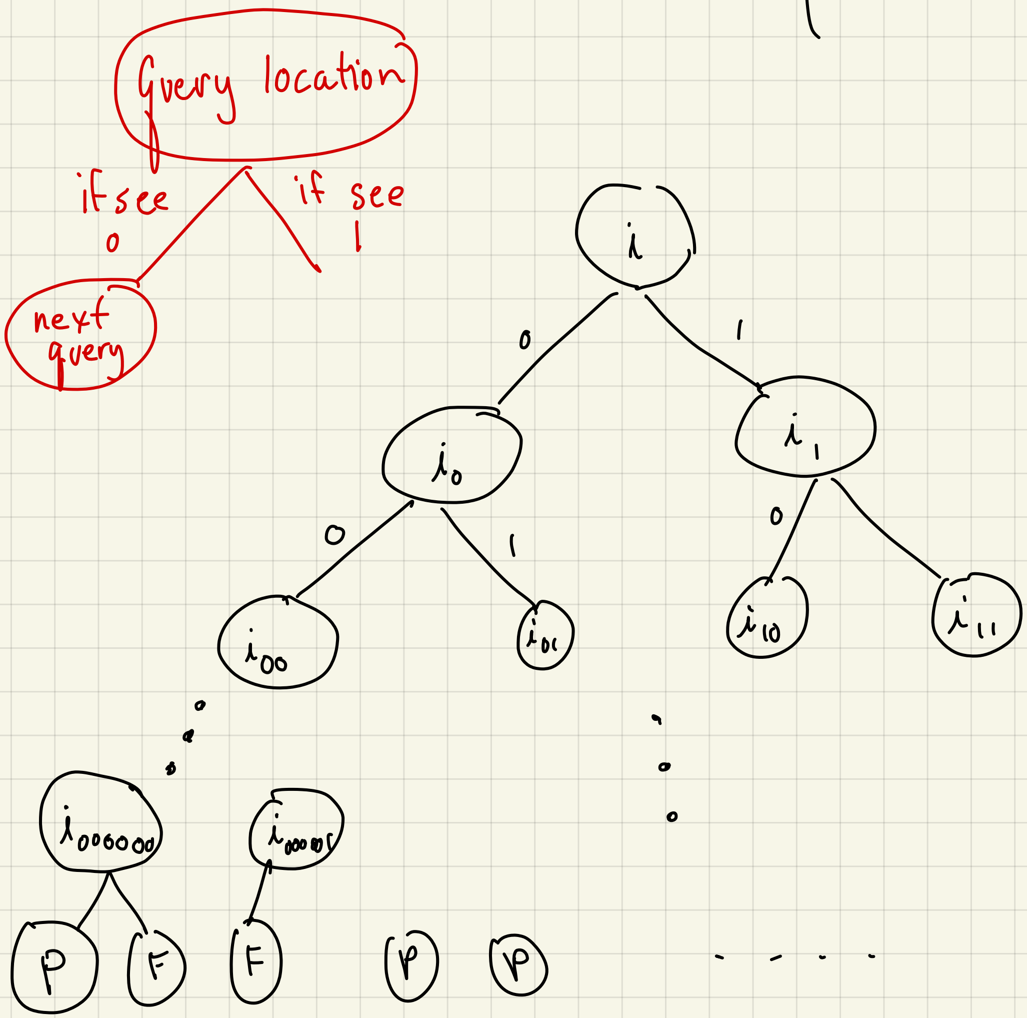
e.g. 0000....0

Bad Distribution:

$$D = \begin{cases} \cdot \ \text{flip coin} \\ \cdot \ \text{if } H \text{ output according to } F \\ \quad \text{else} \quad " \quad " \quad " \quad P \end{cases}$$

Assume deterministic algorithm A s.t. $\begin{cases} \forall\, x \in 2PAL, \quad Pr[A(x) = PASS] \geq 2/3 \\ \& \ \forall\, x \ \varepsilon\text{-far from } 2PAL, \quad Pr[A(x) = FAIL] \geq 2/3 \\ \text{makes } o(\sqrt{n}) \text{ queries} \end{cases}$

describe via query tree:
(for inputs of size n)



hopefully inputs that reach here are supposed to pass

- wlog all leaves have depth t
- $\leq 2^t$ root-leaf paths

We can calculate prob of reaching each leaf given input dist

For each leaf ℓ:

$$E^-(\ell) = \{ \overset{\text{inputs}}{w} \in \{0,1\}^n \text{ s.t. } \underbrace{\text{dist}(w, 2PAL) \geq \varepsilon n}_{w \text{ should FAIL}} \text{ \& } w \text{ reaches leaf } \ell \}$$

$$E^+(\ell) = \{ \overset{\text{inputs}}{w} \in \{0,1\}^n \underbrace{\cap\, 2PAL}_{w \text{ should PASS}} \text{ \& } w \text{ reaches leaf } \ell \}$$

Total error of $\mathcal{A}$ on D:

$$= \sum_{\ell \text{ passing}} \Pr_{w \in D}[w \in \underset{\text{should Fail}}{E^-(\ell)}] + \sum_{\ell \text{ failing}} \Pr_{w \in D}[w \in \underset{\text{should PASS}}{E^+(\ell)}]$$

For each leaf ℓ:

$$E^-(\ell) = \{ \overset{\text{inputs}}{w} \in \{0,1\}^n \text{ s.t. } \underbrace{\text{dist}(w, 2PAL) \geq \varepsilon n}_{w \text{ should FAIL}} \text{ \& } w \text{ reaches leaf } \ell \}$$

$$E^+(\ell) = \{ \overset{\text{inputs}}{w} \in \{0,1\}^n \underbrace{\cap 2PAL}_{w \text{ should PASS}} \text{ \& } w \text{ reaches leaf } \ell \}$$

Total error of A on D:

$$= \sum_{\ell \text{ passing}} \Pr_{w \in D}[\underset{\text{should fail}}{w \in E^-(\ell)}] + \sum_{\ell \text{ failing}} \Pr_{w \in D}[\underset{\text{should PASS}}{w \in E^+(\ell)}]$$

Claim 1 if $t = o(n)$, $\forall \ell$ at depth t

$$\Pr_D[w \in E^-(\ell)] \geq \left(\tfrac{1}{2} - o(1)\right) 2^{-t}$$

} so "Fail" inputs show up at all leaves

Claim 2 if $t = o(\sqrt{n})$, $\forall \ell$ at depth t

$$\Pr_D[w \in E^+(\ell)] \geq \left(\tfrac{1}{2} - o(1)\right) 2^{-t}$$

} so "PASS" inputs show up at all leaves

But each leaf has to choose one label! will be wrong on almost 1/2

Total error of A on D:

$$= \sum_{\ell \text{ passing}} \left(\tfrac{1}{2} - o(1)\right) 2^{-t} + \sum_{\ell \text{ failing}} \left(\tfrac{1}{2} - o(1)\right) 2^{-t} \geq \tfrac{1}{2} - o(1) >> \tfrac{1}{3}$$

$\boxtimes$

Claim 1 if $t = o(n)$, $\forall \ell$ at depth t

$$\Pr_D[w \in E^-(\ell)] \ge \left(\tfrac{1}{2} - o(1)\right) 2^{-t}$$

so "Fail" inputs show up at all leaves

F = random string of distance $\ge \varepsilon n$ from 2PAL

$P = \begin{cases} 1. & \text{pick } k \in_R [\frac{n}{6}+1, \frac{n}{3}] \\ 2. & \text{pick random } v, u \text{ s.t. } |v| = k \quad |u| = \frac{n-k}{2} \\ 3. & \text{output } v v^R u u^R \end{cases}$

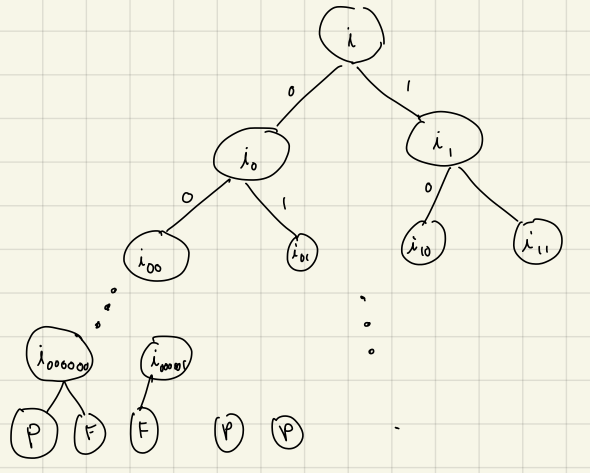
D =
- flip coin
- if H output according to F
  else " " " P

Proof:

Plan:

- F is close to U
- U is uniformly distributed at each leaf (each locn has random bit, so go left/right with equal probability)

$$\Longrightarrow \Pr_{w \in U}[w \in E^-(\ell)] = \frac{2^{n-t}}{2^n} = 2^{-t}$$



But how much can distribution on leaves change using F?

(input size n) $|2PAL_n| \le 2^{\frac{n}{2}} \cdot \frac{n}{2}$ ← choice of x

choice of u, v

$$\#\text{ words at distance } \varepsilon \text{ from } 2PAL_n \le 2^{\frac{n}{2}} \cdot \frac{n}{2} \cdot \sum_{i=0}^{\varepsilon n} \binom{n}{i} \le 2^{n/2 + 2\varepsilon \log(\frac{1}{\varepsilon}) \cdot n}$$

Very few!!

so $\quad E^-(\ell) \geq \underbrace{2^{n-t}}_{\text{\# strings in } \mathcal{U} \text{ reaching } \ell} - \underbrace{2^{\frac{n}{2}+2\varepsilon\log\frac{1}{\varepsilon}n}}_{\text{\# words not in } F} = (1-o(1))2^{n-t}$

- assume $\varepsilon << 1/8$
- t is $o(n)$

so 1st term swamps this

so $\quad \Pr_D[w \in E^-(\ell)] \geq \frac{1}{2}\Pr_F[w \in E^-(\ell)]$

$$= \frac{1}{2}\frac{|E^-(\ell)|}{2^n} \geq \left(\frac{1}{2}-o(1)\right)2^{-t} \qquad \blacksquare$$

Claim 2 if $t = o(\sqrt{n})$, $\forall \ell$ at depth t

$$\Pr_b[w \in E^+(\ell)] \ge \left(\tfrac{1}{2} - o(1)\right)2^{-t}$$

so "PASS" inputs show up at <u>all</u> leaves

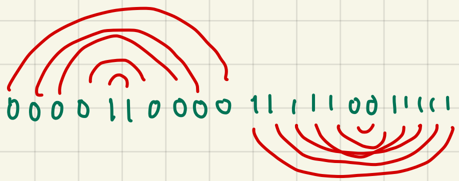
Proof Plan: for every fixed set of $o(\sqrt{n})$ queries, lots of strings in 2PAL follow the path

how many strings agree with leaf ℓ? 2^{n-t}

how many n-bit strings in 2PAL agree with leaf ℓ?

$\ge 2^{\frac{n}{2}-t} - ???$

difficulty:



Fix $k=10$: should see same value at:

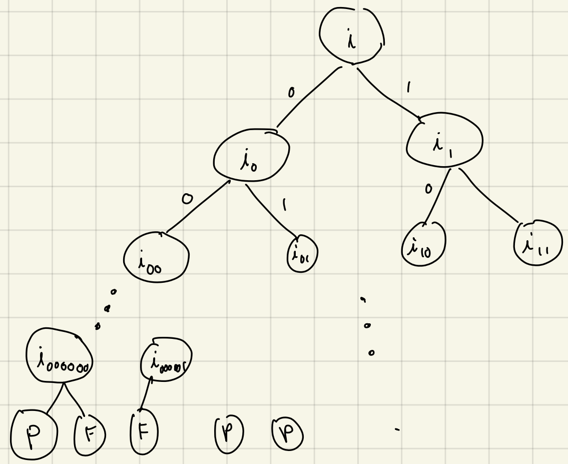
1, 10
2, 9
3, 8
⋮

Lots of "dependencies"

F = random string of distance $\ge \varepsilon n$ from 2PAL

$P = \begin{cases} 1. & \text{pick } k \in_R \left[\frac{n}{6}+1, \frac{n}{3}\right] \\ 2. & \text{pick random } v, u \text{ s.t. } |v| = k \quad |u| = \frac{n-k}{2} \\ 3. & \text{output } vv^R uu^R \end{cases}$

D =
- flip coin
- if H output according to F
  else " " " P

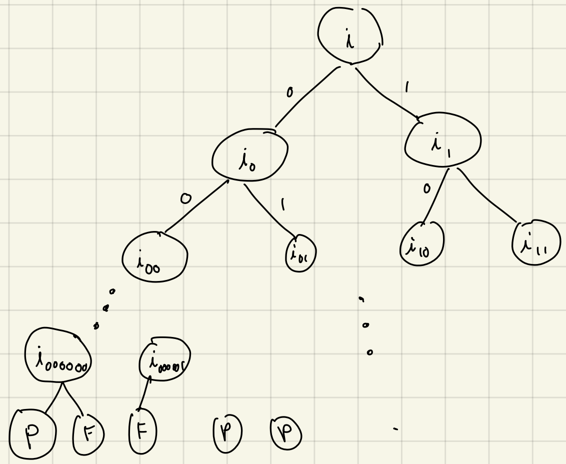


Maybe <u>no</u> string follows path?

but K is picked randomly! in $\left[\frac{n}{6}+1, \dots, \frac{n}{3}\right]$

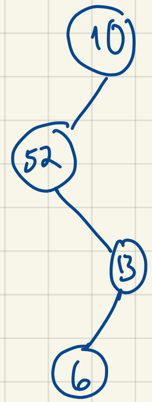
hope: paths that pair up dependent queries for one K will do badly on most others?

$$P = \begin{cases} 1. & \text{pick } K \in_R \left[\frac{n}{6}+1, \frac{n}{3}\right] \\ 2. & \text{pick random } v, u \text{ s.t. } |v| = k \quad |u| = \frac{n-k}{2} \\ 3. & \text{output } vv^R uu^R \end{cases}$$

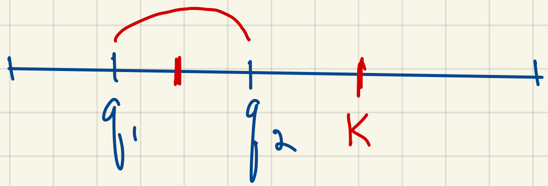


Consider leaf ℓ,

$Q_\ell \leftarrow$ indices queried along way



$\forall$ pair $q_1, q_2 \in Q_\ell$, at most 2 choices of K "pair" them:



only 1 choice in this case

$\Rightarrow$ # choices of K s.t. $\underline{\underline{\text{no pair}}}$ in Q_ℓ $\underbrace{\text{symmetric around } k \text{ or } \frac{n}{2}+k}_{\text{"Good" } K}$

$$\geq \frac{n}{6} - 2\binom{t}{2} = (1 - o(1))\left(\frac{n}{6}\right)$$

Claim 2 if $t = o(\sqrt{n})$, $\forall \ell$ at depth t

$$\Pr_D[w \in E^+(\ell)] \ge \left(\tfrac{1}{2} - o(1)\right)2^{-t}$$

so "PASS" inputs show up at all leaves

Proof

Plan: for every fixed set of $o(\sqrt{n})$ queries, lots of strings in 2PAL follow the path

how many strings agree with leaf ℓ? 2^{n-t}

how many n-bit strings in 2PAL agree with leaf ℓ?

$\ge 2^{\frac{n-t}{2}} - ???$

\# choices of k s.t. no pair in Q_ℓ $\underbrace{\text{symmetric around } k \text{ or } \tfrac{n}{2}+k}_{\text{"Good" } k}$

$$\ge \frac{n}{6} - 2\binom{t}{2} = (1-o(1))\left(\frac{n}{6}\right)$$

So $$\Pr_P[w \in E^+(\ell)] = \sum_w \sum_k \underbrace{\Pr_P[w|k]}_{2^{-n/2}} \cdot \underbrace{\Pr[\text{choose } k]}_{6/n} \cdot 1_{w \in E^+(\ell)}$$

$$\ge \frac{1}{\left(\frac{n}{6}\right)\left(2^{n/2}\right)} \cdot \left[(1-o(1))\frac{n}{6}\right] 2^{\frac{n}{2}-t} = (1-o(1))\,2^{-t}$$

F = random string of distance $\ge \varepsilon n$ from 2PAL

P =
1. pick $k \in_R [\frac{n}{6}+1, \frac{n}{3}]$
2. pick random v, u s.t. $|v| = k$, $|u| = \frac{n-k}{2}$
3. output $vv^R uu^R$

D =
- flip coin
- if H output according to F
  else " " " P