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:

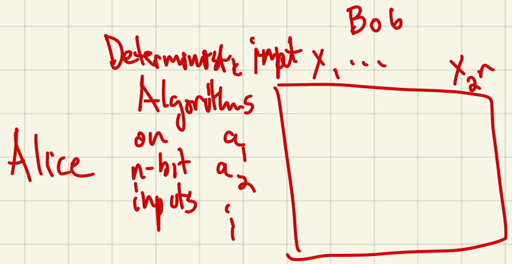
average case lower bound

$\Downarrow$

randomized worst case l.b.

Yao's ~~Principle~~ Thm.: 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) (see Wikipedia)



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 (includes random bits)

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

$\Rightarrow$ if want to show l.b. need only show a "bad" distribution on inputs that is "hard" for any deterministic algorithm

Example application of Yao's method:

$2PAL = \{ w \mid w \text{ is } w = vv^R uu^R \}$ Concatenation of 2 palindromes

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

$\boxed{V_1 V_2 \cdots V_K \mid V_K V_{K-1} \cdots V_2 V_1 \mid u_1 u_2 \cdots \mid \quad u_2 u_1}$

Note that testing $PAL = \{ w \mid w = vv^R \}$ is "easy"

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

Can test 2PAL in $O(\sqrt{n})$ time

Can you do better?

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

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

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

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

Pf. Plan: give distribution on inputs that is hard for all deterministic 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 $6 | n$)

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

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

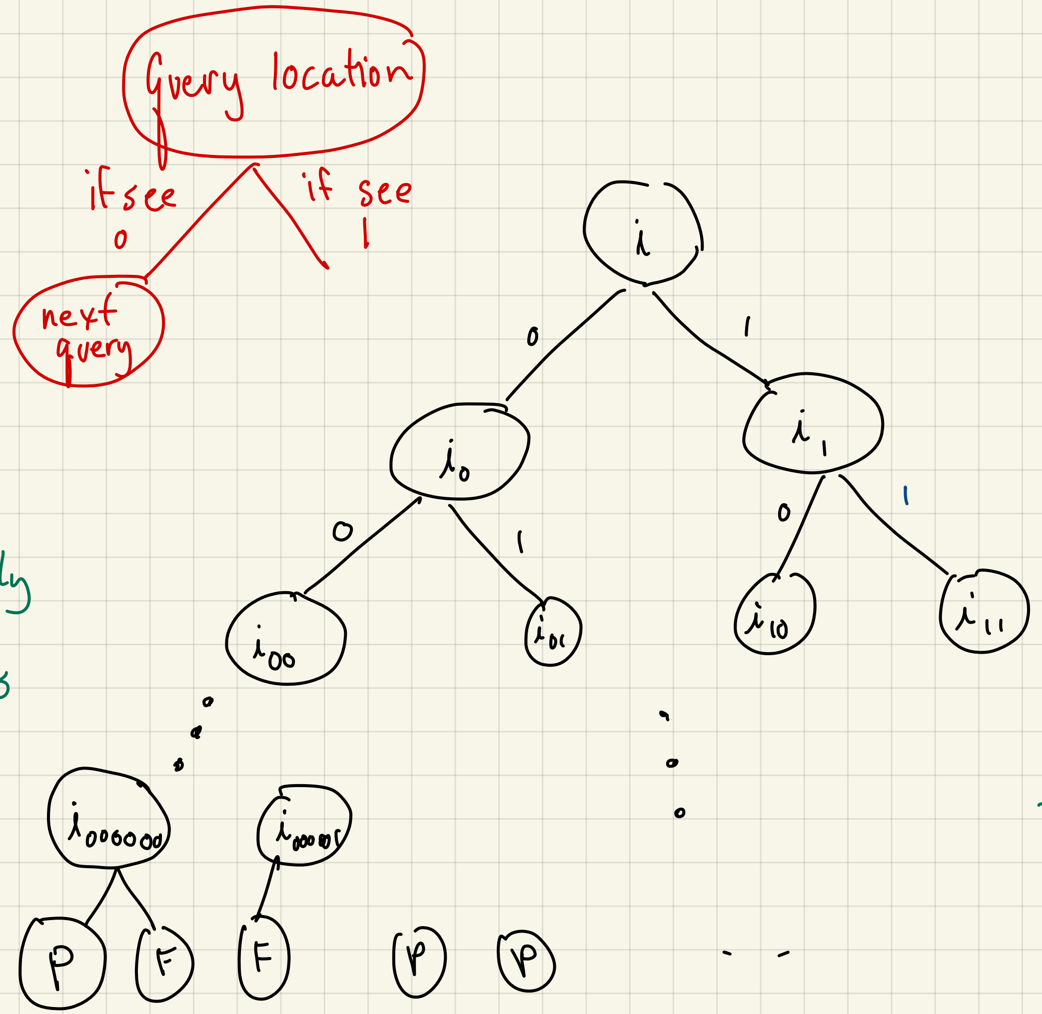
e.g. $111 \cdots 1$

Bad Distribution:

$$D = \begin{cases} \text{flip coin} \\ H: \text{ output according to } F \\ T: \text{ output according to } P \end{cases}$$

Assume deterministic algorithm A

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



assume no "repeats" along path

- depth of decisiontree is t
- wlog assume all leaves have depth t (complete binary tree)
- 2^t root-leaf paths

each input follows exactly one branch
reaches leaf, which is hopefully labelled by correct answer

we can calculate prob of reaching each leaf given (input dist F_p)

Suppose $w \in_R \{0,1\}^n$: $\Pr[w \text{ reaches leaf } \ell] = 2^{-t}$

For each leaf ℓ:

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

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

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

$$= \sum_{\ell \ \underbrace{\text{passing}}_{\text{output of leaf}}} \Pr_{w \in \mathcal{D}} [w \in \underbrace{E^-(\ell)}_{\text{correct answer is Fail}}] + \sum_{\ell \ \text{failing}} \Pr_{w \in \mathcal{D}} [w \in \underbrace{E^+(\ell)}_{\text{should PASS}}]$$

For each leaf ℓ:

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

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

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

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

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

$$\Pr_D[w \in E^-(\ell)] \geq \underbrace{\left(\tfrac{1}{2} - o(1)\right)}_{\text{almost } 1/2 \text{ prob that random inputs reach } \ell} 2^{-t}$$

} 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 a label so will be wrong on almost 1/2 inputs that reach it

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

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

$\blacksquare$

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

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

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

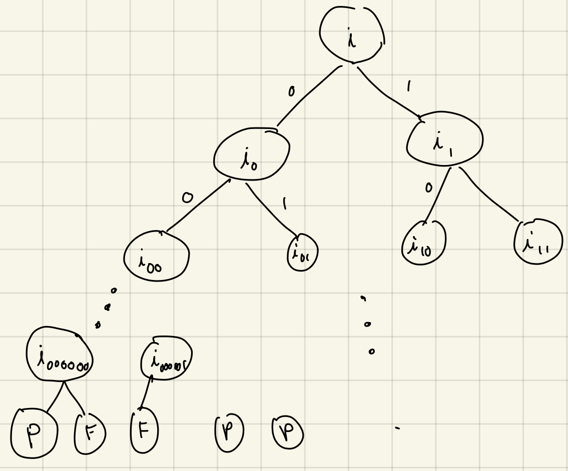
$$D = \begin{cases} \cdot & \text{flip coin} \\ \cdot & \text{if } H \text{ output according to } F \\ & \text{else} \quad \text{" " " } P \end{cases}$$

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)

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



But how much can distribution on leaves change using F?

$$|2PAL_n| \leq \underbrace{\frac{n}{2}}_{\text{choice of } i} \cdot \underbrace{2^{n/2}}_{\text{choice of } u, v}$$

$$\#\text{ words at dist } \leq \varepsilon n \text{ from 2PAL} \leq \left(2^{n/2} \cdot \frac{n}{2}\right) \sum_{i=0}^{\varepsilon n} \binom{n}{i} \leq 2^{n/2 + 2\varepsilon \log \frac{1}{\varepsilon} \cdot n} \quad \text{very few}$$

so $E^-(\ell) \geq 2^{n-t} - 2^{\frac{n}{2} + 2\varepsilon \log \frac{1}{\varepsilon} \cdot n} = (1 - o(1))\, 2^{n-t}$ since $t << \frac{n}{2}$

\# strings in u that reach ℓ

\# words not in F

assume $\varepsilon << 1/8$, t is $o(n)$ } so 1st term swamps this

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

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

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

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

so "PASS" inputs show up at all leaves

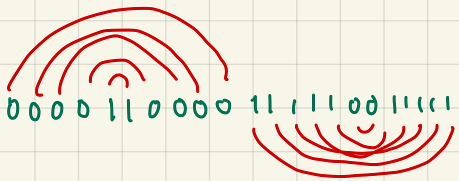
Proof Plan: for every fixed set of $o(\sqrt{n})$ queries, lots of strings in 2PAL follow the path

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

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

$\geq 2^{\frac{n}{2} - ??} - ??$

difficulty:



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

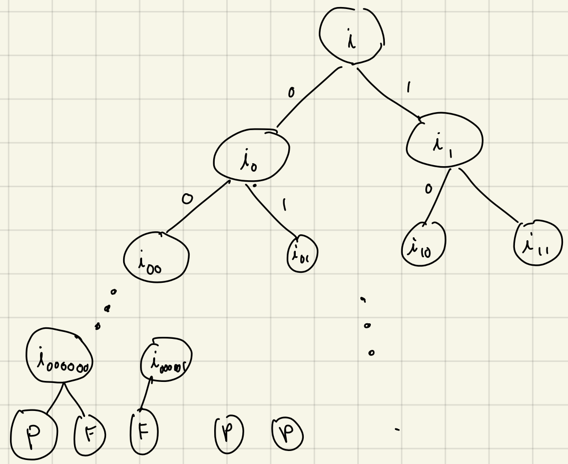
1, 10
2, 9
3, 8
⋮

Lots of dependencies

F = random string of distance $\geq \varepsilon n$ 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

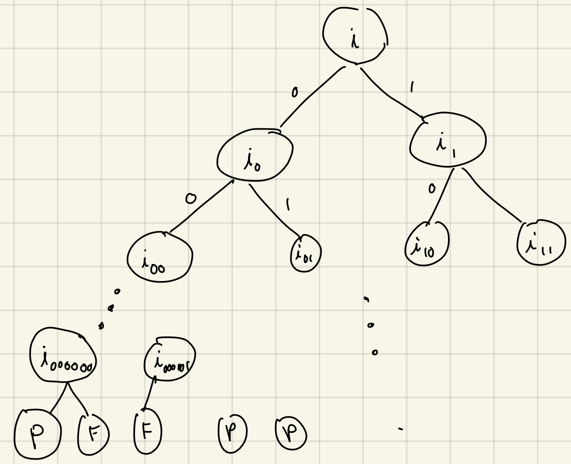


Maybe no string in 2PAL follows the path?

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

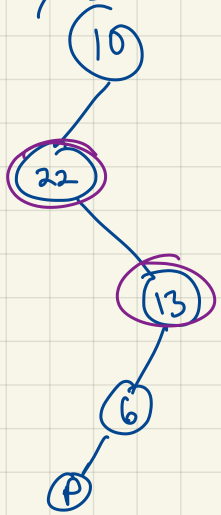
$P = \begin{cases} 1. & \text{pick } k \in_R [\frac{n}{6}+1, \frac{n}{3}] \\ 2. & \text{pick random } v, u \text{ s.t. } |v| = k \quad |u| = \frac{n-k}{2} \\ 3. & \text{output } vv^R uu^R \end{cases}$

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



Consider leaf ℓ,

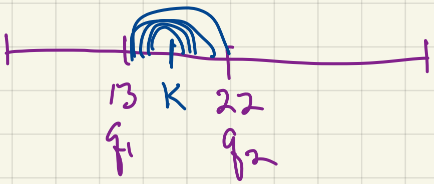
$Q_\ell = \{10, 22, 13, 6\}$



$Q_\ell \leftarrow$ indices queried along way

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

1?



if p picked k that pairs $q_1 + q_2$ then all bets off

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

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

1?

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

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?

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

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

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

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

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

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

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

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