LECTURE 15: COMMUNICATION COMPLEXITY

General communication problem:
System needs to perform some task
Depends on info distributed among different parts of system (processors / parties / players)
Thus, players need to <u>communicate</u> to solve task

Simplified abstract model [Yao '79]
- Two parties / players (Alice & Bob)
- Each player gets fixed part of input
- Only care about communication (so players computationally unbounded)
- Compute prespecified function of input

This model seems to hit a sweet spot
- Very clean and simple, and so can be analyzed
- But yet general enough to be applicable in many different scenarios

Examples of applications (don't have time to go into details):
- VLSI circuit design
- data structures
- streaming algorithms
- Boolean circuit lower bounds
- Property testing
- Proof complexity
- And more...

X, Y, Z arbitrary finite sets
$f: X \times Y \to Z$ arbitrary function
Alice gets $x \in X$, Bob gets $y \in Y$
Want to compute $f(x,y) \Rightarrow$ need to communicate
Use fixed protocol P (depending only on f)
for sending bits to each other

Protocol P must specify
- is communication terminated?
- if so, what is the answer $f(x,y)$?
- if not, who sends the next bit? [Depends only on bits communicated so far.]
- what is the next bit sent? [Depends on communication so far + private input of Alice / Bob]

Only interested in communication
Ignore internal computations of Alice & Bob
i.e., they have unlimited computational powers

Cost of P on (x,y) = # bits communicated
Cost of P = worst case, i.e., $\max_{(x,y)}\{\text{cost of } P \text{ on } (x,y)\}$

Communication complexity of f =
cost of <u>best</u> protocol for f

[In general, will have functions
$$f_n : X_n \times Y_n \to Z_n$$
for $n \to \infty$ and interested in asymptotic analysis. This is usually suppressed from notation.]

<u>DEF 1</u> A protocol P over domain $\mathcal{X} \times \mathcal{Y}$ with range $\mathcal{Z}$ is a binary tree.

Each internal node v labelled by

$a_v : \mathcal{X} \to \{0,1\}$ (if Alice speaks)

or $b_v : \mathcal{Y} \to \{0,1\}$ (if Bob speaks)

Each leaf ℓ is labelled by some $z \in \mathcal{Z}$

Value of P on (x,y) = element z in leaf reached when we walk left if a_v / b_v gives 0 and right if a_v / b_v gives 1.

Cost of P on (x,y) = length of path

Cost of P = max length over all (x,y)

P computes f if value of P on (x,y) is $f(x,y)$.

<u>DEF 2</u> The (deterministic) communication complexity of $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ is the min cost of any protocol P computing f.

Notation: $D(f)$

<u>Many possible variants</u> of this set-up. Here are some:

(1) Function v.s. relation

Let x, y be (sub)sets of some universe $\mathcal{U}$

Function $f(x,y) = [x \cap y \neq \emptyset] = \begin{cases} 1 & \text{if } x \cap y \neq \emptyset \\ 0 & \text{otherwise} \end{cases}$

Relation: Find $w \in x \cap y$

[Focus on functions; perhaps discuss relations in pset 4]

Notation $[\text{condition}] = \begin{cases} 1 & \text{if condition true} \\ 0 & \text{if condition false} \end{cases}$

(2) # players 2 or more

[Focus exclusively on 2 players.
3 or more players important, but much more challenging.]

(3) Must Alice & Bob be deterministic?
Can they flip coins?
Can they share quantum entanglement?
[Today deterministic. Next time random.
We'll do no quantum.]

Since players have unbounded computational powers, a trivial solution is that one player sends his/her input and the other announces the answer

<u>PROP 3</u> For any $f: \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ it holds that

$$D(f) \leq \log_2 |\mathcal{X}| + \log_2 |\mathcal{Z}|$$
$$D(f) \leq \log_2 |\mathcal{Y}| + \log_2 |\mathcal{Z}|$$

<u>EX 4</u> Let $x, y \subseteq \{1, 2, \ldots, n\}$

(Can represent as bit strings $\{0,1\}^n$
$x_i = 1 \iff i \in x$
Switch between these two interpretations freely)

$MED(x,y)$ = median of multiset $x \cup y$
If $t = 2k$ elements, say median is kth smallest.

(a) $D(MED) \leq n + \log_2 n$

Follows immediately from Prop 3

(b) $D(\mathrm{MED}) = O(\log^2 n)$

Idea: binary search
Maintain interval $[i,j]$ s.t. median $\in [i,j]$
At beginning $i=1$, $j=n$
Look at $k = (i+j)/2$
Alice sends # elements in x less than k
and # elements $\geq k$ $O(\log n)$ bits
Now Bob knows if median $\leq k$ or $> k$
Bob sends this info to Alice 1 bit
Both update i,j on their own
Then repeat.
Interval is halved at each stage
$\rightarrow$ $\log_2 n$ stages
$O(\log n)$ communication per stage

__Obviously__ careful attention to details needed in order to get the binary search right, but this is not the point here so we ignore it.

__EX 5__ $x, y \in \{0,1\}^n$

$$EQ(x,y) = [x=y] = \begin{cases} 1 & \text{if } x=y \\ 0 & \text{otherwise} \end{cases}$$

Trivial protocol uses $n+1$ bits.
Is it possible to do better?
Or can we prove a lower bound

(c) Actually $D(\mathrm{MED}) = O(\log n)$, but this protocol is a bit tricky (but fun!)
Won't have time to do it

Let's focus on lower bounds

$$\mathrm{Range}(f) = \{ z \in \mathcal{Z} \mid \exists (x,y) \in \mathcal{X} \times \mathcal{Y} \text{ s.t. } f(x,y) = z \}$$

<u>PROP 6</u> $$D(f) \geq \log_2 |\mathrm{Range}(f)|$$

<u>Proof</u> Protocol = binary tree.
At termination, answer know to both Alice & Bob.
Leaves labelled by $z \in \mathcal{Z}$
At least $m = |\mathrm{Range}(f)|$ distinct leaves
$\Rightarrow$ height of binary tree $\geq \log_2 m$. ☐

<u>COR 7</u> $D(\mathrm{MED}) = \Omega(\log n)$ (and this is right).

<u>But</u> communication complexity mostly focuses on Boolean functions $\mathcal{Z} = \{0,1\}$
Then Prop 6 useless...

[Non-Boolean case also interesting, but most techniques extend from Boolean case to non-Boolean case.]

To prove lower bounds, view protocol as a way of partitioning $\mathcal{X} \times \mathcal{Y}$ into sets S so that for each $(x,y) \in S$ the <u>same communication</u> is sent

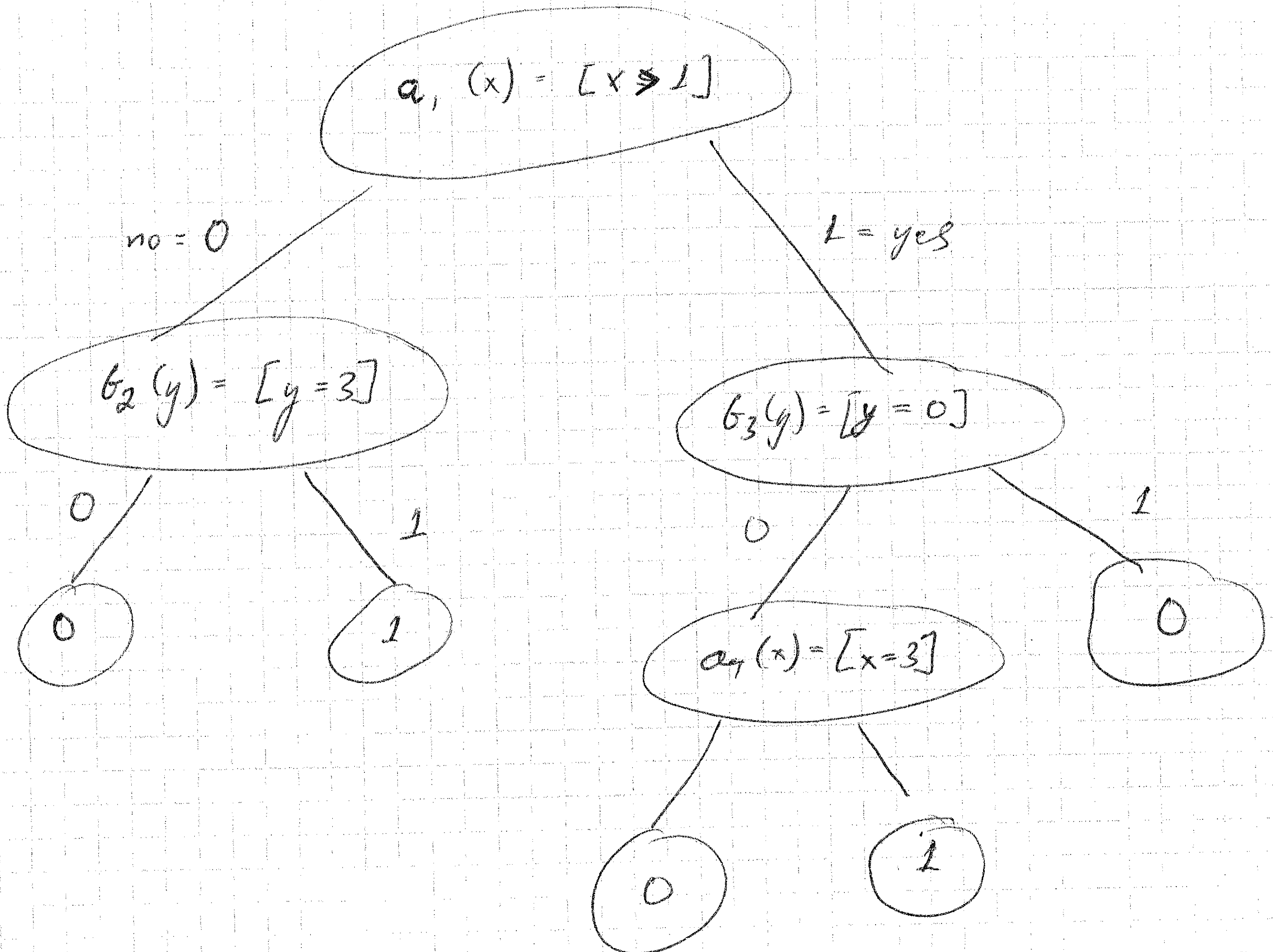
Then analyze structure of such sets and partitions

For an illustration, see function and protocol on next page.

Ex 8

$x, y \in \{0, 1, 2, 3\}$

If $x \le 1$ then $[y = 3]$
else if $x = 2$ then 0
else $[y \neq 0]$



Function table

x \ y	00	01	10	11
00	0	0	0	1
01	0	0	0	1
10	0	0	0	0
11	0	1	1	1

Can also view this as matrix M_f with $M_{x,y} = f(x,y)$

For P protocol, v node in protocol tree, let

$$R_v = \{ (x,y) \mid \text{protocol reaches node } v \text{ on input } (x,y) \}$$

If $\mathcal{L} = \{ \text{leaves of } P \}$ then $\{R_\ell\}_{\ell \in \mathcal{L}}$ is a partition of $\mathcal{X} \times \mathcal{Y}$.

<u>DEF 9</u> A <u>(COMBINATORIAL) RECTANGLE</u> in $\mathcal{X} \times \mathcal{Y}$ is a subset $R \subseteq \mathcal{X} \times \mathcal{Y}$ such that $R = A \times B$ for some $A \subseteq \mathcal{X}$, $B \subseteq \mathcal{Y}$.

<u>PROP 10</u> $R \subseteq \mathcal{X} \times \mathcal{Y}$ is a rectangle iff $(x_1, y_2) \in R$ and $(x_2, y_2) \in R \Rightarrow (x_1, y_2) \in R$

<u>Proof</u> Exercise.

<u>PROP 11</u> For every protocol P and every protocol node v, R_v is a rectangle (in particular, this holds for leaves).

<u>Proof</u> Use Proposition 10. Suppose that (x_1, y_1), $(x_2, y_2) \in R_v$ and show $(x_1, y_2) \in R_v$

Proof by induction. True for root r of protocol tree ($R_r = \mathcal{X} \times \mathcal{Y}$).

Suppose w child of v; $(x_1, y_1), (x_2, y_2) \in R_w$

Then $(x_1, y_1), (x_2, y_2) \in R_v$ so by IH $(x_1, y_2) \in R_v$

If Alice speaks at v, she only sees transcript and x_1, and so will go to w for both (x_1, y_1) and (x_1, y_2).

If Bob speaks, transcript looks the same for (x_1, y_2) and (x_2, y_2), so Bob moves to same node on both. $\boxed{2}$

Intuitively, every time Alice speaks the scenarios for (x_1, y_1) and (x_1, y_2) look exactly the same, so she says the same thing by definition, and analogously for Bob.

$R \subseteq \mathcal{X} \times \mathcal{Y}$ is MONOCHROMATIC if f takes the same ~~value~~ $f(x,y)$ on all $(x,y) \in R$.

If ℓ leaf of protocol, then R_ℓ is monochromatic by definition. Hence:

<u>LEMMA 12</u> Any protocol $\mathcal{P}$ for $f: \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ induces a partition of $\mathcal{X} \times \mathcal{Y}$ into monochromatic rectangles.

[# rectangles = # leaves in protocol tree]

<u>COR 13</u> If <u>any</u> partition of $\mathcal{X} \times \mathcal{Y}$ into monochromatic rectangles requires $\geq t$ rectangles, then $D(f) \geq \log_2 t$

<u>Proof</u> Same as for Prop 6

Need $\geq t$ leaves $\Rightarrow$ tree height $\geq \log_2 t$ $\boxed{4}$

Look at Ex 8 again

	00	01	10	11
00	0	0	0	1
01	0	0	0	1
10	0	0	0	0
11	0	1	1	1

Note that combinatorial rectangles need <u>not</u> be contiguous geometric rectangles (but it's easier to draw such rectangles in examples).

FOOLING SETS [Yao '79, Lipton & Sedgewick '81]

Exhibit large set of input pairs s.t. no two pairs (x,y), (x',y') can be in the same monochromatic rectangle
$\Rightarrow$ # monochromatic rectangles large
$\Rightarrow$ CC lower bound

<u>DEF 14</u> $S \subseteq X \times Y$ is a FOOLING SET for $f : X \times Y \to \{0,1\}$ if $\exists\, z \in \{0,1\}$ s.t.

(a) $\forall (x,y) \in S \quad f(x,y) = z$

(b) $\forall (x_1,y_1),(x_2,y_2) \in S$, $(x_1,y_1) \neq (x_2,y_2)$, either $f(x_1,y_2) \neq z$ or $f(x_2,y_1) \neq z$.

<u>LEMMA 15</u> If f has fooling set S then $D(f) \geq \log_2 |S|$.

<u>Proof</u> Any $\mathcal{P}$ partitions $X \times Y \supseteq S$ into monochromatic rectangles. We claim no rectangle can contain more than one $(x,y) \in S$. Hence #rectangles $|S|$. Apply Cor 13

Proof of Claim: Suppose R contains $(x_1, y_1), (x_2, y_2) \in S$. Then by Prop 10 R also contains (x_1, y_2) and (x_2, y_1).
But by definition of fooling set f does not have the same value on these 4 inputs.
Contradiction.

Slight optimization: Do argument in Lem 15 twice, once for 0-rectangles and once for 1-rectangles.

<u>Ex 16</u> Consider $EQ(x,y) = [x = y]$ for $x, y \in \{0,1\}^n$.

$S = \{ (\alpha, \alpha) \mid \alpha \in \{0,1\}^n \}$
is a fooling set of size 2^n for $z = 1$

$EQ(\alpha, \alpha) = 1 \qquad EQ(\alpha, \beta) = 0$ for $\alpha \neq \beta$.

$S' = \{ (0^n, 1^n), (1^n, 0^n) \}$ is a fooling set of size 2 for $z = 0$.

Lem 15 yields $D(EQ) > n$.
Hence $D(EQ) = n + 1$.

<u>Ex 17</u> $x, y \subseteq \{1, 2, \dots, n\}$ (i.e. $\in \{0,1\}^n$)
$DISJ(x,y) = [x \cap y = \emptyset]$

A fooling set of size 2^n is $S = \{ (A, \bar{A}) \mid A \subseteq \{1, 2, \dots, n\} \}$
For $A \neq B$ either $A \cap \bar{B}$ or $\bar{A} \cap B$ is non-empty.
Hence $D(DISJ) \geq n$.
Count also 0-rectangles $\Rightarrow D(DISJ) > n$
Trivial upper bound $D(DISJ) \leq n + 1$ — tight!

Set disjointness DISJ is a very important problem in comm cplx. — the "NP-complete" problem of comm cplx.

Question $D(EQ) = n+1$

$D(DISJ) = n+1$

What happens if we allow randomness? Do these problems become any easier?

Fooling set technique special case of following lemma

LEMMA 18 Let μ be a probability distribution on $X \times Y$. If any monochromatic rectangle R has measure $\mu(R) \leq \delta$, then $D(f) \geq \log_2 (1/\delta)$.

Proof $\mu(X \times Y) = 1$ by def. Hence at least $1/\delta$ rectangles in any monochromatic partition. $\boxed{\Sigma}$

For fooling set S, set

$$\mu((x,y)) = \frac{1}{|S|} \quad \text{for } (x,y) \in S$$

$$\mu((x,y)) = 0 \quad \text{for } (x,y) \notin S.$$

Another application of Lem 18:

If all monochromatic rectangles are small, say of size $\leq k$, then # rectangles $\geq |X||Y|/k$ and hence

$$D(f) \geq \log |X| + \log |Y| - \log k.$$

Can also prove lower bounds
by studying the rank of the matrix M_f

Can study deterministic communication
with 3 players or more.

We won't do this, but there is some
more material in Ch 13 in Arora-Barak
(optional).