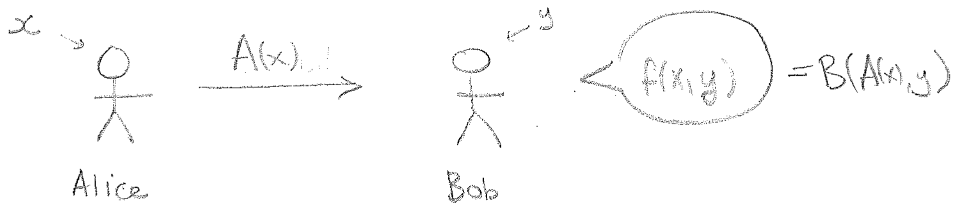
1

# (ONE WAY) Communication Complexity

**Agenda**

- Deterministic 1-way communication complexity
- Relationship to streaming algs
- Randomized CC:
  - $\vec{R}(\mathrm{DISJ}) \geq \Omega(n)$
  - $\vec{R}(\text{GAP-HAMMING}) = \Omega(n)$




Alice is given $x \in X$
Bob $y \in Y$
Alice sends one message, $A(x)$, to Bob
Bob must output $f(x,y) = B(A(x), y)$

Def] The Deterministic, one-way communication complexity of $f$, $\vec{D}(f)$, is the smallest $s \in \mathbb{N}$ s.t. $\exists$ $A: X \to \{0,1\}^s$, $B: \{0,1\}^s \times Y \to \{0,1\}$ s.t. $\forall x \in X, y \in Y$, $B(A(x), y) = f(x,y)$.

such an $(A,B)$ is a "deterministic one-way protocol"

Examples!

- Equality $\mathrm{EQ}(x,y) = \begin{cases} 1 & x = y \\ 0 & x \neq y \end{cases}$, $x, y \in \{0,1\}^n$

  $\vec{D}(\mathrm{EQ}) = n$.

  Indeed, otherwise $\exists\ x, x' \in \{0,1\}^n$ s.t. $A(x) = A(x')$.

- Disjointness $\mathrm{DISJ}(x,y) = \begin{cases} 1 & x \cap y = \phi \\ 0 & x \cap y \neq \phi \end{cases}$ $x, y \in \{0,1\}^n$, $\subseteq [n]$

  $\vec{D}(\mathrm{DISJ}) = n$. (Same arg).

- Count $\mathrm{COUNT}(x,y) = \sum_i (x_i + y_i)$ (over $\mathbb{R}$) $x, y \in \{0,1\}^n$

  $\vec{D}(\mathrm{COUNT}) = \log_2(n)$

  Upper bound: Alice sends $\sum_i x_i$

  lower bound: else, might be two different counts for $\sum_i x_i$ that map to the same thing.

## (2) Application: Streaming algs.

A streaming alg $\mathcal{A}$ w/ space $s$

- takes one pass over a stream of input $a_0, a_1, \ldots, a_{m-1}$ w/ $a_i \in \mathcal{U}$
- Uses $\leq s$ space
- outputs something at the end of the stream. (it doesn't know $m$ in advance)

<u>Example</u>: frequency moments.

$$F_0 = \#\text{distinct elts in } \{a_0, \ldots, a_{m-1}\}$$

$$F_\infty = \max_{u \in \mathcal{U}} |\{i : a_i = u\}|$$

<u>Claim</u> Any deterministic streaming alg that approximates $F_\infty$ (i.e., $\forall \vec{a}$, $\mathcal{A}(\vec{a}) = (1 \pm 1/4) F_\infty(\vec{a})$) $(*)$ uses at least space $|\mathcal{U}|$.

<u>proof</u> Suppose $\mathcal{A}$ is an alg w/ $(*)$.

Here's a protocol for DISJ:

- Alice starts running $\mathcal{A}$ w/ $a_0, a_1, \ldots$ = elements of her input $x \subseteq [n]$
- Alice sends the memory of $\mathcal{A}$ to Bob.
- Bob finishes running $\mathcal{A}$ on his input.
- Bob outputs "disjoint" if $\mathcal{A}(x, y) \leq 5/4$

This works:

If $x \cap y \neq \phi$, then $F_\infty(\vec{a}) = 2$, so $\mathcal{A}(x,y) \geq 2\left(\frac{3}{4}\right) = 3/2$.

If $x \cap y = \phi$, $F_\infty(\vec{a}) \leq 1$, so $\mathcal{A}(x,y) \leq 1 \cdot (5/4) < 3/2$

- Similar proof shows you need $\Omega(|\mathcal{U}|)$ to compute $F_0$ (deterministically, exactly).
- Also true: $\Omega(|\mathcal{U}|)$ needed to compute $F_0$ approximately, deterministically.

(3) Natural question: What about Randomized protocols?

<u>Thm</u> There exists a randomized streaming alg $\mathcal{A}$ so that $\forall \vec{a}$,

$$\mathbb{P}\{ \mathcal{A}(\vec{a}) \in (1 \pm \epsilon) F_0 \} \geq 9/10,$$

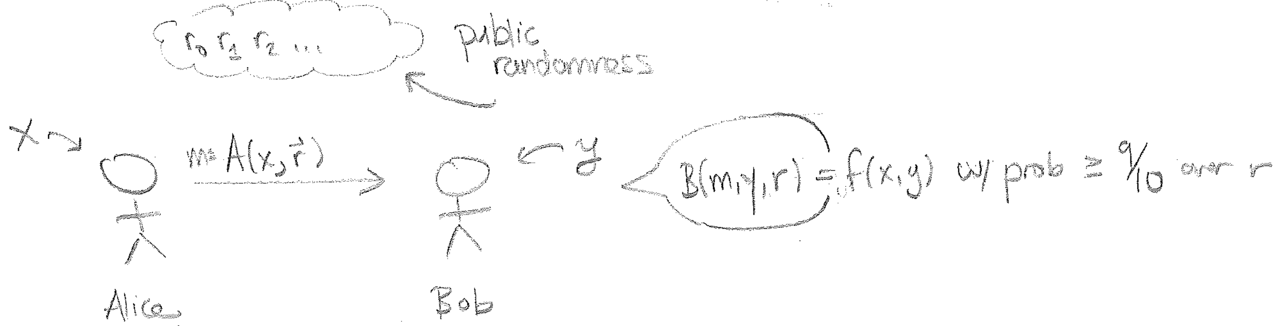
space is $O\left( \frac{\log(|\mathcal{U}|) + \log(m)}{\epsilon^2} \right)$.

m = length of stream

pf (HW).

Is the same true for $F_\infty$?

<u>Randomized One Way Communication Complexity.</u>




<u>Def</u> The randomized one-way communication complexity $\vec{R}(f)$ is the smallest $s$ s.t.

$\exists\ A: X \times \{0,1\}^* \to \{0,1\}^s$

$\exists\ B: \{0,1\}^s \times Y \times \{0,1\}^* \to \{0,1\}$ s.t.

such an A, B are a "randomized 1-way protocol"

$$\mathbb{P}_{r \sim \mathcal{U}} \{ B(A(x,r), y, r) = f(x,y) \} \geq 9/10.$$

<u>Thm</u> $\vec{R}(\mathrm{DISJ}) = \Omega(n)$

<u>Cor</u> Any randomized streaming alg that approximates $F_\infty$ w/ prob $\geq 9/10$ requires space $\Omega(|\mathcal{U}|)$.

(and, any randomized streaming alg that computes $F_0$ exactly w/ prob $\geq 9/10$ requires $\Omega(|\mathcal{U}|)$ space)

(slightly different arg)

④

Proof of thm: Distributional Complexity

<u>Lemma</u> (Yao). Let $D$ be a distribution on $\mathbb{X} \times \mathbb{Y}$, $\varepsilon \in (0, 1/2)$.
Suppose that every deterministic 1-way protocol $P$ w/

$$\mathbb{P}_{x,y \sim D} \left( P(x,y) \neq f(x,y) \right) \leq 1/10$$

has communication cost at least $t$.
Then $\vec{R}(f) \geq t$.

That is, we can push randomness in $P$ into randomness over the inputs; our lower-bound task is now to find a hard distribution.

Proof. Let $R$ be a randomized protocol, w/ communication cost $\not\leq t$.
$R$ can be written as a distribution over deterministic protocols: for each choice of $r$, we get a deterministic protocol, $P_r$.

For each $r$,

$$\mathbb{P}_{(x,y) \sim D} \left[ P_r(x,y) \neq f(x,y) \right] > 1/10 \qquad \text{by assm.}$$

So.

$$\mathbb{P}_{(x,y) \sim D,\ r \sim \mathcal{U}} \left[ R(x,y,r) \neq f(x,y) \right] > 1/10 \qquad \text{avg'ing over } r$$

So $\exists\, (x,y)$ s.t. $\mathbb{P}_{r \sim \mathcal{U}} \left[ R(x,y,r) \neq f(x,y) \right] > 1/10$.

Then $\vec{R}(f) \geq t$. ∎

5 Recall we're trying to bound $\vec{R}(\text{DISJ}) \geq \Omega(n)$.
We'll do this by reducing to another problem.

The indexing problem:

$$X = \{0,1\}^n, \quad Y = [n], \quad \text{INDEX}(x,i) = x_i.$$

Thm $\vec{R}(\text{INDEX}) = \Omega(n)$.

Notice that this will imply $\vec{R}(\text{DISJ}) = \Omega(n)$. Indeed, say that we had a protocol for DISJ. Here's a protocol for INDEX, w/ same cost + error prob:

On input $(x, i)$:

- Run protocol for DISJ on $(x, e_i)$
- Return 0 if protocol says "disjoint", else return 1.

If $x \cap e_i = \phi$, then $x_i = 0$
$x \cap e_i \neq \phi$, then $x_i = 1$

So now let's prove that $\vec{R}(\text{INDEX}) = \Omega(n)$.

Pf. Use distributional complexity. Let's try the uniform dist on $\{0,1\}^n \times [n]$. Turns out we'll get lucky and this will work.

Let $\delta < 1$ sufficiently small (TBD), $n$ sufficiently big.
Say there's a (deterministic) protocol for indexing w/ space $\leq \delta \cdot n$, so that

$$\mathbb{P}_{\substack{x \sim \{0,1\}^n \\ i \in [n]}} \left[ P(x,i) \neq x_i \right] \leq 1/10$$

Let $B(x) = (P(x,1), P(x,2), \ldots, P(x,n))$

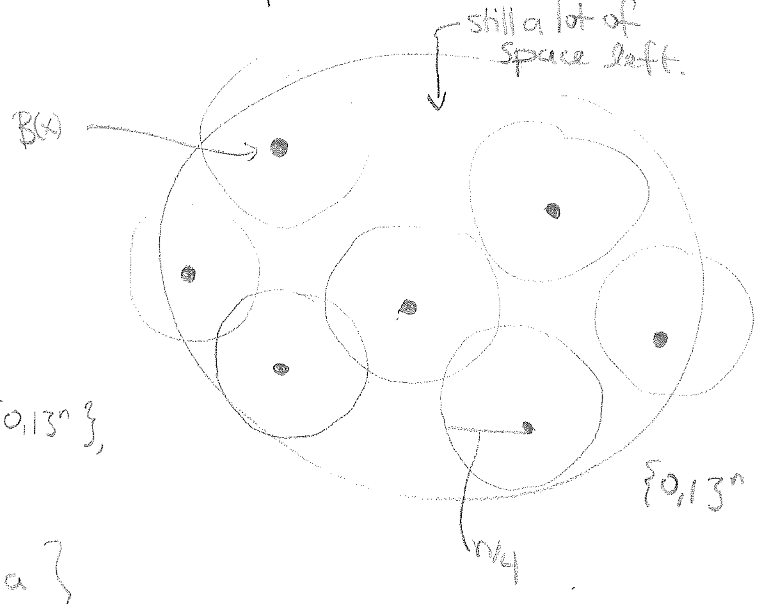
what Bob outputs on $A(x), 2$ in the deterministic protocol P.

So for a fixed $x$,

$$\mathbb{P}_{i \in [n]} \{P(x,i) \neq x_i\} = \frac{d_H(x, B(x))}{n} \quad \longleftarrow \text{hamming distance}$$

(6) But, $B(x)$ can only take $2^{\delta n}$ possible values.

Claim: At least $2^{n-1}$ $x$'s have $d_H(x, B(x)) \geq n/4$.




Proof. Let $\mathcal{C} = \{ B(x) : x \in \{0,1\}^n \}$, so $|\mathcal{C}| \leq 2^{\delta n}$.

{For those who know lingo, $\mathcal{C}$ is a rate-$\delta$ code in $\{0,1\}^n$.}

Notice $d_H(x, B(x)) < n/4 \Rightarrow \exists\, c \in \mathcal{C},\ d_H(x,c) < n/4$

$$\#\{x : d(x, B(x)) < n/4\} \leq \#\{x : \exists c \in \mathcal{C},\ d_H(x,c) < n/4\}$$
$$\leq 2^{\delta n} \cdot \left(\begin{array}{c}\text{number of } z \in \{0,1\}^n \text{ w/} \\ \text{hamming wt} < n/4\end{array}\right)$$
$$\leq 2^{\delta n} \cdot 2^{0.9n}$$
$$\leq 2^{0.95n} < 2^{n-1}$$

this is a very crude bound:

if $\delta < 0.05$ is small enough ∎

$$\sum_{j=1}^{n/4} \binom{n}{j} \leq n \cdot \binom{n}{n/4} \leq n\,(4e)^{n/4} \leq n \cdot 2^{\lg(4e)\, n/4} = n \cdot 2^{0.86n}$$

Say $x$ is "good" if $d_H(x, B(x)) < n/4$.

So, at least half the $x$ are bad.

Now we can finish the thm.

$$\mathbb{P}_{\substack{x \sim \{0,1\}^n \\ i \in [n]}} \left[ P(x,i) \neq x_i \right] = \overbrace{\mathbb{P}_x(x \text{ is good}) \cdot \mathbb{P}_{x,i}\{ P(x,i) \neq x_i \mid x \text{ is good} \}}^{\geq 0}$$

$$+ \underbrace{\mathbb{P}_x(x \text{ is bad})}_{\geq 1/2} \; \underbrace{\mathbb{P}_{x,i}\{ P(x,i) \neq x_i \mid x \text{ is bad} \}}_{\substack{= \frac{1}{n} d_H(x, B(x)) \text{ for each } x \\ \geq 1/4}}$$

$$\geq 1/8 > 1/10.$$

We just showed space $\leq \delta n \Rightarrow$ failure prob $> 1/10$.

So every deterministic protocol w/ success probability $\geq 9/10$ must use space $> \delta n$.

By Yao's lemma, $\vec{R}(\text{INDEX}) \geq \delta n$. ∎

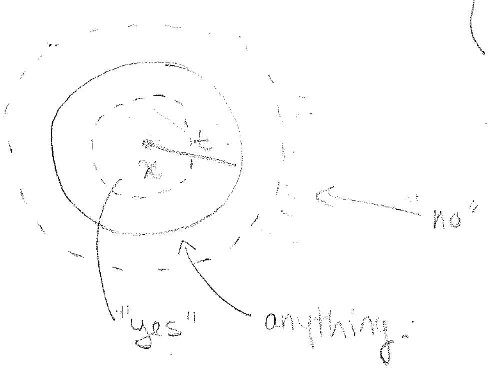
As we saw, this implies $\vec{R}(\text{DISJ}) \geq \delta n$, which implies that any randomized alg for appx $F_\infty$ needs space $\Omega(|U|)$.

8

Another problem: Gap-Hamming.
w/ a slicker reduction

Alice has $x \in \{0,1\}^n$
Bob has $y \in \{0,1\}^n$

$$\text{GAP-HAMMING}(t) = \begin{cases} 1 & d_H(x,y) < t - c\sqrt{n} \\ 0 & d_H(x,y) > t + c\sqrt{n} \end{cases}$$




c is some small constant

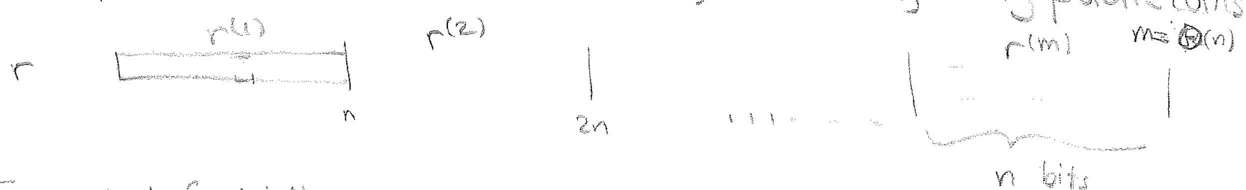
Why do we care? This is kind of like disjoint~~ness~~, but more expressive. We'll be able to use it to prove fancier lower bds on streaming algs.

Consider GH = GAP-HAMMING($n/2$).

$\boxed{\text{THM}}$ $\vec{R}(GH) = \Omega(n)$

$\boxed{\text{pf}}$ Randomized reduction from INDEX.

Alice has $x \in \{0,1\}^n$. Bob has $i \in [n]$. Say n is odd.
They will generate an instance $(a, b)$ of GH, using public coins.




To get their first bits:

Alice:
if $d_H(r^{(1)}, x) > n/2$, $a_1 = 1$
if $d_H(r^{(1)}, x) < n/2$, $a_1 = 0$

Bob:
$b_1 = r_i^{(1)}$.

→

9

Notice: $a_1$ and $b_1$ are correlated!

Indeed, condition on the $n-1$ bits of $r^{(1)}$ other than $r_i^{(1)}$.

CASE 1. $x, r$ agree on $\not\leq \frac{n-1}{2}$ or $\not\geq \frac{n-1}{2}$ bits off of $i$

Then $a_1$ is already determined, so
$\mathbb{P}(a_1 = b_1) = \mathbb{P}(a_1 = r_i) = 1/2$ by independence of $r_i$

CASE 2. $x, r$ agree on exactly $\frac{n-1}{2}$ bits, off of $i$

Then $x_i = 1 \Rightarrow a_1 = b_1$
$x_i = 0 \Rightarrow a_1 \neq b_1$

if $r_i = 0$, $d_H(r^{(1)}, x) < n/2$, $a_1 = 0$
if $r_i = 1$, $d_H(r^{(1)}, x) \geq n/2$, $a_1 = 1$

Now, $$\mathbb{P}_r(\text{case } 2) = \frac{\binom{n-1}{(n-1)/2}}{2^n} \underset{\substack{\text{Stirling's} \\ \text{appx}}}{\approx} \frac{c' 2^n/\sqrt{n}}{2^n} = c'/\sqrt{n}$$

so $$\mathbb{P}\{a_1 = b_1\} = \underbrace{\mathbb{P}(\text{case } 1)}_{1 - c'/\sqrt{n}} \underbrace{\mathbb{P}(a_1 = b_1 \mid \text{case } 1)}_{1/2} + \underbrace{\mathbb{P}(\text{case } 2)}_{c'/\sqrt{n}} \underbrace{\mathbb{P}(a_1 = b_1 \mid \text{case } 2)}_{\underset{x_i=1}{1} \text{ or } \underset{x_i=0}{0}}$$

$$= \begin{cases} 1/2 + \frac{c'}{2\sqrt{n}} & x_i = 1 \\ 1/2 - \frac{c'}{2\sqrt{n}} & x_i = 0 \end{cases}$$

Now do this $m$ times, $m = Cn$.

$$\mathbb{E}\, d_H(a, b) = \begin{cases} m\left(1/2 + \frac{c'}{2\sqrt{n}}\right) = m/2 + C''\sqrt{m} & x_i = 1 \\ m/2 - C''\sqrt{m} & x_i = 0 \end{cases}$$

Chernoff: w/ prob $\geq 9/10$

$$d_H(a, b) \begin{cases} \geq m/2 + C'''\sqrt{m} & x_i = 1 \\ \leq m/2 - C'''\sqrt{m} & x_i = 0 \end{cases}$$

So it's a Gap-hamming instance!

Thus, if there's a randomized protocol for Gap-Hamming w/ success prob $9/10$ and communication $o(n)$, then there's one for INDEX w/ $\underbrace{\text{success prob } 8/10}_{\text{hence } \exists \text{ one w/ prob } 9/10 \text{ by repeating}}$, comm $o(n)$ and we saw that can't happen ↯