Lower Bounds for Randomized Communication Complexity and Streaming Algorithms

Purdue University | Young-San Lin
CS 590 SLA Lecture 20 | April 1, Spring 2021

This scribe note uses a lot of notations from the last one. One might want to read the last one before reading this one.

1 Preliminaries

We consider communication protocols with public random coin, which is equivalent to considering distributions over deterministic protocols. The cost of a randomized communication protocol is the worst-case cost over inputs (x, y) and randomness.

Lemma 1.1 (Yao 83'). Let D be a distribution over (x, y) and let $\varepsilon \in (0, 1/2)$. Suppose every deterministic one-way protocol P s.t.

$$\Pr_{(x,y)\sim D}[P \text{ is wrong}] \leq \varepsilon$$

has communication complexity at least k, then every randomized algorithm with public random coin with error at most ε on every input has communication cost at least k.

Proof. Let R be a randomized algorithm of communication cost smaller than k which is wrong w.p. at most ε on every input (x, y). Regard R as a distribution over deterministic protocols $P_1, P_2, ..., P_T$ with cost smaller than k, then

$$\Pr_{(x,y)\sim D}[P_i \text{ is wrong}] > \varepsilon \quad \forall i \in [T].$$

We also have that

$$\Pr_{(x,y)\sim D,R}[R \text{ is wrong}] > \varepsilon \quad \forall i \in [T]$$

where the randomness is based on the randomized algorithm R.

There exists (x, y) s.t.

$$\Pr_{i\in[T]}[P_i \text{ is wrong}] > \varepsilon \quad \forall i \in [T],$$

which implies that R is wrong w.p. at least ε on the specific input (x, y), a contradiction. □

The converse statement also holds by a proof using von Neumann minimax duality[1]. The proof is omitted here.

[1] This can also be viewed as a two-person zero-sum game where one person picks the distribution over the input to maximize the cost while the other picks the distribution over the deterministic algorithms to minimize the cost.

Lemma 1.2. Let $\varepsilon \in (0, 1/2)$. Suppose every randomized protocol R with public random coin s.t.

$$\Pr[R \text{ is wrong}] \leq \varepsilon$$

on every input has communication complexity at least k, then there exists a distribution over (x, y), s.t. every deterministic algorithm with error at most ε has communication cost at least k.

Now we can state Yao's lemma in a simpler form.

Theorem 1.3. $R_\delta(f) = \max_{\mu \text{ over independent } (x,y)} D_\delta^\mu(f)$.

Intuitively, the worst-case error of a random protocal is the error of the best distribution over deterministic algorithms against the distribution over the input, which is equivalent to that of the worst distribution over the input against the distribution over deterministic algorithms.

We note Yao's lemma can be used not only in communication complexity, but also in any general setting where the cost comes from the randomized algorithm (or the distribution over deterministic algorithms) verses the distribution of the input.

2 An Application of Yao's Lemma

Theorem 2.1 (Kremer et al. 99')**.** $\vec{R}(INDEX_n) = \Omega(n)$.

Proof. By Yao's lemma, it suffices to find a hard distribution over (x, y) s.t. every deterministic algorithm uses at least cn bits of communication.

We show that the following distribution is hard for every deterministic algorithm. We pick x uniformly at random (u.a.r) from $\{0, 1\}^n$ and i u.a.r from $[n]$ where x and i are picked independently. Let's say (x, i) is drawn from U in this scenario.

Claim 2.2. Every deterministic protocol that uses at most cn bits is wrong w.p. at least 1/8.

Suppose Alice reads x and sends z to Bob. Bob computes $(B(z, 1), B(z, 2), ..., B(z, n))$ based on what he receives. The protocol P is correct on (x, i) if $B(z, i) = x_i$. We want to show that

$$\Pr_{(x,i)\sim U}[P \text{ is wrong on } (x, i)] > \frac{1}{8}$$

if P uses fewer than cn bits.

We observe that

$$\Pr_i[P \text{ is wrong on } (x, i) \mid x, z] = d_H(x, B(z))$$

where d_H denotes the relative Hamming distance. This is because if we fix x and z, i is picked u.a.r.

Let B be the set of all possible vectors computed by Bob, i.e., for all z, $B(z) \in B$. There are at most 2^{cn} many z's.

Definition 2.3. x is *good* if there exists $b \in B$ s.t. $d_H(x, b) \leq 1/4$. Otherwise, x is *bad*.

Now we have that

$$\begin{aligned}
&\Pr_{(x,i)\sim U}[P \text{ is wrong on } (x, i)] \\
&= \Pr_i[P \text{ is wrong } \mid x \text{ is good}] \Pr[x \text{ is good}] + \Pr_i[P \text{ is wrong } \mid x \text{ is bad}] \Pr[x \text{ is bad}] \\
&\geq 0 + \frac{1}{4} \Pr[x \text{ is bad}].
\end{aligned}$$

Therefore, it suffices to show that $\Pr[x \text{ is bad}] \geq 1/2$. This is equivalent to showing that w.p. at least $1/2$, x does not lie in the union of balls formed by $b \in B$ as centers with radius $n/4$ where $|B| \leq 2^{cn}$.

We show that the number of bad x's is at least 2^{n-1}. We observe that

$$|Ball(b, \frac{n}{4})| = \sum_{i=0}^{n/4} \binom{n}{i} \leq n \binom{n}{n/4} \leq n \left(\frac{ne}{n/4}\right)^{n/4} \leq n(4e)^{n/4} = n2^{n \log 4e/4} \leq 2^{0.862n},$$

$$|\bigcup_i Ball(b_i, \frac{n}{4})| \leq 2^{cn} \cdot 2^{0.862n} = 2^{(c+0.862)n} < 2^{n-1}$$

for large n. The probability of picking a good x is smaller than $1/2$. □

3 Other Reductions

Theorem 3.1. $\vec{R}(DISJ_n) \geq \Omega(n)$.

Sketch. We reduce from INDEX to DISJ. We want to transform an input (x, i) for INDEX to an input (x, ℓ_i) for DISJ, and show that $x_i = 0$ iff the answer for (x, ℓ_i) is YES. □

Theorem 3.2 (Kalganasundaran and Schnitger 92')**.** $R(DISJ_n) = \Omega(n)$.

Theorem 3.3. Every randomized p-pass streaming algorithm for connectedness must use $\Omega(n/p)$ space.

Sketch. We reduce from DISJ. Given the input (x, y) for DISJ, we create a graph $G = (\{s, t\} \cup [n], E)$ consisting of Alices vertex s, Bob's vertex t, and one layer of vertices $[n]$. $(s, i) \in E$ iff $x_i = 0$ and $(t, i) \in E$ iff $y_i = 0$. We observe that G is connected iff $x \cap y = \emptyset$. We have p passes, and for each path, we can use n/p space to track connectivity. □