CS-448 Sublinear Algorithms for Big Data Analysis April 16, 2025

Lecture 9

Lecturer: Michael Kapralov

In several exercise sessions so far we have proved streaming lower bounds for deterministic algorithms. In this lecture we start designing tools for proving lower bounds for randomized algorithms.

1 Communication complexity

Lower bounds for randomized algorithms are typically proved via reductions from appropriately defined communication problems.

We will study two party communications problems, where Alice holds input $X \in \{0,1\}^n$, Bob holds input $Y \in \{0,1\}^n$, and they want to compute a function $f(X,Y)$. We will study one-way communication problems, where Alice compresses her input X into a message m, and sends the message to Bob. Bob then outputs the answer based on m and his input Y. We now outline the relation to streaming algorithms. Suppose that there exists a small space streaming algorithm ALG for computing $f(X,Y)$ when X and Y are given in a stream. ALG yields a communication efficient protocol as follows: Alice feeds ALG her part of the input, i.e. X, then communicates the state of the memory of the algorithm to Bob, who finishes the execution of ALG and outputs the answer. Thus, if we can prove a lower bound of s bits on the one-way communication complexity of f, a lower bound of s bits on the space complexity of ALG follows immediately.

We will consider several communication settings in what follows.

Let $D(f)$ denote the minimum communication complexity of a *deterministic* communication protocol for computing f. Let $R_\delta^{pub}(f)$ denote the minimum communication complexity of a randomized communication protocol for computing f with error probability at most δ on every input, where Alice and Bob have access to a source of shared randomness. Let $R_\delta^{pri}(f)$ denote the minimum communication complexity of a randomized communication protocol for computing f with error probability at most δ on every input, where Alice and Bob only have private randomness. Let $D_{\mu,\delta}(f)$ denote the distributional complexity of computing f with error probability at most δ over inputs X,Y drawn from the distribution μ.

2 The INDEX problem

Alice has $x \in \{0,1\}^n$ and Bob is given $i \in [n]$. Then, the goal is to compute $f(x,i) = x_i$ on Bob's end with a single message m from Alice. Recall that $R_\delta^{pub,\to}(f)$ stands for the public coin one-way communication complexity of computing a function $f(x,y)$ with error probability at most δ on every input: Alice holds x, Bob holds y, they share a source of random bits and Alice sends a single message to Bob, after which he must output the correct answer with probability at least $1-\delta$ on every fixed pair of inputs.

Claim 1

$$R_\delta^{pub\to}(INDEX) \geq (1 - H_2(\delta))n$$

where $H_2(\delta) = \delta \log_2 \frac{1}{\delta} + (1-\delta)\log_2 \frac{1}{1-\delta}$ *is the binary entropy at* δ*.*

2.1 Information theory crash course

Let X and Y be discrete random variables. Define

- *Entropy:* $H(X) = \sum_x p(x) \log_2 \frac{1}{p(x)} = \mathbf{E}_X[\log_2 \frac{1}{p(x)}]$
- *Joint entropy:* $H(X,Y) = \sum_{(x,y)} p(x,y) \log_2 \frac{1}{p(x,y)} = \mathbf{E}_{X,Y}[\log_2 \frac{1}{p(x,y)}]$
- *Conditional entropy:* $H(X|Y) = \sum_y p(y) H(X|Y=y) = \mathbf{E}_Y[H(X|Y=y)]$
- *Mutual information:* $I(X;Y) = H(X) - H(X|Y)$

Lemma 2 *The following relations hold:*

- *Chain rule for entropy:* $H(X,Y) = H(X) + H(Y|X)$
- *Chain rule for mutual information:* $I(X;Y,Z) = I(X;Z) + I(X;Y|Z)$
- *Entropy subadditivity:* $H(X,Y) \leq H(X) + H(Y)$
- *Conditioning does not increase entropy:* $H(X|Y) \leq H(X)$
- $H(X) \leq \log_2(|supp(X)|)$
- *Data processing inequality: for any function* f *one has* $H(f(X)) \leq H(X)$

2.1.1 Fano's inequality

Theorem 3 *Let* X *and* Y *be discrete random variables and* g *an estimator (based on* Y*) of* X *such that* $\Pr[g(Y) \neq X] = \delta$*. Then,*

$$H(X|Y) \leq H_2(\delta) + \delta \log_2(|supp(X)| - 1)$$

Intuition suggests that as estimator $g(Y)$ for X gets better (e.g. lower error probability), then Y reduces the uncertainty (entropy) about X. Fano's inequality makes this intuition quantitative. Lastly, note that for binary X the second term in the right hand side of the inequality is 0.

Equipped with the information theoretic claims above, we can now give a proof of Claim 1:

Proof of Claim 1: Let X denote the length n vector that Alice holds, and let $X \sim UNIF(\{0,1\})^n$. Let the size of the message that Alice sends be s (we can assume without loss of generality that Alice always sends messages of the same length), and let M be the message. First note that

$$R_\delta^{pub,\to}(M) \geq H(M) \geq I(M;X),$$

The first inequality follows since Alice sends s bits, so $|\text{supp}(M)| \leq 2^s$, and thus $H(M) \leq s$. The second inequality follows from the definition of mutual information and nonnegativity of entropy: $I(M;X) = H(M) - H(M|X) \leq H(M)$.

By correctness of the protocol we know that **for any** x **and** i Bob correctly guesses x_i with probability at least $1-\delta$ (over randomness in Alice's message), i.e., for every i there exists g_i such that $\Pr_M[g_i(M(x)) \neq x_i] \leq \delta$. Letting for any $i \in [n]$ $X_{<i}$ denote the vector $(X_1, \ldots, X_{i-1})$, we get

$$\begin{aligned}
I(X;M) &= \sum_{i=1}^{n} I(X_i; M|X_{<i}) \quad \text{(chain rule for mutual information)} \\
&= \sum_{i=1}^{n} H(X_i|X_{<i}) - H(X_i|M, X_{<i}) \\
&\geq \sum_{i=1}^{n} H(X_i) - H(X_i|M) \;\; (X_i \text{ are iid, and conditioning does not increase entropy}) \\
&= \sum_{i=1}^{n} (1 - H(X_i|M)) \quad (X_i \text{ is a uniform binary r.v.}) \\
&\geq (1 - H_2(\delta))n, \quad \text{(by correctness of INDEX } \exists g_i : \Pr[g_i(M) \neq X_i] \leq \delta \text{ and Fano's inequality)}
\end{aligned}$$

as required. ■