CS 229r: Algorithms for Big Data Fall 2013

Lecture 8 — Sept. 26, 2013

Prof. Jelani Nelson *Scribe: Anudhyan Boral*

1 Overview

In this lecture we begin a new topic: space lower bounds for streaming algorithms. Our main tool is going to be communication complexity. Some examples of such space lower bounds are:

- We need $\Omega(\varepsilon^{-2} + \log n)$ bits of space for the F_0 problem in the streaming model.
- Computing the exact median deterministically requires $\Omega(n)$ space.
- 2-approximation of F_p requires $\Omega(n^{1-\frac{2}{p}})$ bits. (Hence for $p > 2$ we must use at least polynomial space instead of polylogarithmic).

2 Communication Complexity

Consider a (co-operative) game between Alice and Bob. Alice is given an element x from some large set X. Bob is given an element y from some large set Y. Alice and Bob want to compute a function $f : X \times Y \to \{0, 1\}$ by collaborating with each other. They want to minimize the number of bits of communication between them.

Depending on the number of rounds allowed, Alice and Bob take turns passing messages to each other. First, Alice sends Bob a message m_0. Then, Bob sends Alice a message m_1, and so on. The communication game with r rounds involves the sending of r messages.

There are several variants of the communication game. They differ, for instance, in the person having to decide the value of $f(x, y)$. It could be Alice, or Bob or even a third party agent observing all the communication.

The main observation which allows us to use the theory of communication complexity is: *One-way communication lower bounds imply streaming space lower bounds.*

3 Different Types of Communication Complexity

3.1 Deterministic Communication Complexity

We denote by $D(f)$ the minimum number of bits required by a deterministic communication protocol that always correctly computes f. That is, Alice and Bob behave deterministically.

3.2 Randomized Communication Complexity

$R_\delta^{\text{pri}}(f)$ is the minimum number of bits required to obtain f with success probability at least $1-\delta$ where Alice and Bob both have access to private random coins.

$R_\delta^{\text{pub}}(f)$ is the same as $R_\delta^{\text{pri}}(f)$ except that the random coins are public. (Imagine there is an infinite random string written in the sky, where both Alice and Bob can see it.)

3.3 Distributional Communication Complexity

$D_{\mu,\delta}(f)$ is the *distributional communication complexity*, the number of bits required to compute f when the inputs are drawn from a fixed distribution of inputs μ and the success probability required is $1-\delta$.

It is easy to see that $D(f) \geq R_\delta^{\text{pri}}(f) \geq R_\delta^{\text{pub}}(f) \geq D_{\mu,\delta}(f)$. (The last inequality is shown by invoking Yao's principle.) For randomized streaming algorithms the most relevant model is usually $R_\delta^{\text{pri}}(f)$. But for many examples lower-bounding $R_\delta^{\text{pub}}(f)$ is easier.

The book *Communication Complexity* by Kushilevitz and Nisan [8] is an excellent reference for Communication Complexity.

4 Exact deterministic F_0 requires $\Omega(n)$ bits

To prove the lower bound, we want to reduce some hard communication function to the property we want to compute.

Consider the function $EQ : \{0,1\}^n \to \{0,1\}^n$ defined as: $EQ(x,y) = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{if otherwise} \end{cases}$

Claim 1. $D(EQ) = \Omega(n)$

Proof. The proof is obvious by the pigeonhole principle. □

The reduction goes as follows: Suppose we are given a streaming algorithm $\mathcal{A}$ which computes F_0 using $o(n)$ space. Using her input $x \in \{0,1\}^n$, Alice constructs the stream $s_x = \{i : x_i = 1\}$. Alice streams s_x through $\mathcal{A}$ and sends Bob the memory content m_0 of $\mathcal{A}$. Using m_0, Bob can figure out the string x. Initializing $\mathcal{A}$ with the memory content m_0, if feeding it $\{i\}$ changes the value of F_0 then $x_i = 0$; otherwise $x_i = 1$. Hence, as Bob knows x as well as his own input y, he can easily compute $EQ(x,y)$.

5 The Indexing Problem

The communication problem of Indexing is as follows:

- Alice gets $x \in \{0,1\}^n$.

- Bob gets $j \in [n]$.
- $f(x, j) = x_j$.

Claim 2. *(Indexing lower bound)* $R^{pub}_{\delta}(f) \geq (1-H_2(\delta)n$, *where* $H_2(\delta) = -\delta \log_2(\delta) - (1-\delta)\log_2(1-\delta)$.

Before proving this claim we introduce some basic information theory definitions and properties in the next few sections.

6 Information Theory

Let X take values in some domain $\mathcal{X}$.

Definition 3. $H(X) = \sum_{x \in \mathcal{X}} p_x \log_2(p_x)$, *where* $p_x = \mathbb{P}[X = x]$.

Definition 4. $H(X, Y) = -\sum_{x \in \mathcal{X}, y \in \mathcal{Y}} p_{x,y} \log_2(p_{x,y})$, *where* $p_x = \mathbb{P}[X = x \text{ and } Y = y]$.

Definition 5. *(conditional entropy)* $H(X|Y) = \mathbb{E}_{y \in Y} H(X|Y = y)$.

Definition 6. *(mutual information)* $I(X;Y) = H(X) - H(X|Y)$

Intuitively, the mutual information measures the amount of randomness left in X, if Y is known. It can be proved (using simple manipulation) that $I(X;Y) = H(Y) - H(Y|X)$.

7 Basic Properties of Entropy Functions

1. *Chain rule* $H(X, Y) = H(X) + H(Y|X)$
2. *Subadditivity* $H(X, Y) \leq H(X) + H(Y)$ and equality holds iff X and Y are independent.
3. $H(X|Y) \leq H(X)$
4. $H(X) \leq \log(|X|)$.
5. *Fano's Inequality* Suppose there is a deterministic function g such that $\mathbb{P}[g(Y) \neq X) \leq \delta$, then $H(X|Y) \leq H(X|g(Y)) \leq H_2(\delta) + \delta \log_2(|\mathcal{X}| - 1)$.

8 Back to Indexing Lower Bound

Suppose Bob gets the transcript of communication Π and his input $J \in [n]$. We lower-bound the distributional complexity. Choose the 'hardest' distribution - which is the uniform distribution for J. And suppose Bob is able to predict x_j with probability $\geq \delta$.

Using Fano's inequality, we get,

$$\begin{aligned}
H_2(\delta) &\geq H(X_J|\Pi, J) \\
&= \sum_{j=1}^{n} \mathbb{P}[J=j] \cdot H(X_J|J=j,\Pi) \\
&= \frac{1}{n}\sum_{j=1}^{n} H(X_j|\Pi) \\
&\geq \frac{1}{n}\sum_{j=1}^{n} H(X_j|\Pi, X_1, \cdots, X_{j-1}) \text{ (conditioning on more variables can only reduce entropy)} \\
&= \frac{1}{n}(H(X_1, \cdots, X_n, \Pi) - H(X_1, X_2 \cdots, X_{n-1} \cdots + H(x_1, \Pi) - H(\Pi)) \text{ (using chain rule)} \\
&= \frac{1}{n}(H(X_1, \cdots, X_n) - H(\Pi)) \text{ (by telescoping sums)} \\
&= 1 - \frac{1}{n}H(\Pi) \\
&\geq 1 - \frac{1}{n}|\Pi| \text{ (since } H(\Pi) \leq |\Pi| \text{)}
\end{aligned}$$

9 Probabilistic Exact Median Lower Bound

Claim 7. *The exact median problem where success probability is $\geq 1-\delta$ requires at least $(1 - H_2(\delta))n)$ bits of space, if all integers are in $[n]$ and the stream length is $2n-1$.*

Proof. The proof is via reduction from the Indexing problem.

Alice constructs the virtual stream $s_A = [2 + x_1, 4 + x_2, \cdots, 2i + x_i, \cdots]$, and Bob constructs the virtual stream $s_B = s_1 s_2$ where $s_1 = [0, 0, \cdots 0]$ ($n - j$ copies) and $s_2 = [2n + 2, 2n + 2, \cdots 2n + 2]$ ($j - 1$ copies).

Observe that the median of the concatenation of s_A and s_B is $2j + x_j$. If Alice sends Bob the memory content of the algorithm, then Bob can run the algorithm on the concatenated streams, and hence figure out the value of x_j. (Recall that he already knows j).

□

10 The F_0 Lower Bound

We know that computing F_0 with error ε requires $\Omega(\varepsilon^{-2} + \log n)$ bits of space. The $\log n$ comes from [1]. Here we will show the $\Omega(1/\varepsilon^2)$ bound. In our proof, we will use reduce the Gap Hamming problem.

The Gap Hamming Problem: Alice is given $x \in \{0,1\}^N$; Bob is given $y \in \{0,1\}^N$. $f = \Delta(x, y)$, where $\Delta(x, y)$ is the Hamming distance between x and y. That is, the number of bit positions where x and y differ. The error allowed is upto $\pm\sqrt{N}$.

Theorem 8. $R^{pub}_{1/3}$ *(Gap Hamming)* $= \Omega(N)$.

The proof for the above theorem with one-way communication by a reduction from indexing can be found in [7]. The first proof of the optimal one way communication lower bound for Gap Hamming was in [9].

Claim 9. *Computing F_0 with error ε requires $\Omega(\varepsilon^{-2})$ bits. (As long as $\varepsilon^{-2} \leq n$)*

Proof. We reduce from Gap Hamming. Set $N = c\varepsilon^{-2}$. Alice and Bob get $x, y \in \{0,1\}^N$ respectively. Observe that $2F_0 = w(x) + w(y) + \Delta(x, y)$. (Here $w(x)$ denotes the number of 1's in x).

Alice sends Bob $w(x)$, the weight (number of 1's) of x, along with the memory content of the algorithm after the stream s_x. Bob estimates $\Delta(x, y) = 2F_0 - w(x) - w(y)$. □

11 Disjointness Problem and the F_p Lower Bound

We now introduce the communication problem (t-player) Disjointness which is useful for proving the space lower bounds for F_p.

First, we introduce a generic t-player communication game. There are t Alice's denoted by $A_1, A_2, \cdot A_t$. Each A_i is given an input x_i. A_1 sends a message to A_2, A_2 sends one to A_3 and so on. One round consists of $t-1$ message sendings and after that A_t has to report the value of $f(x_1, x_2, \cdots, x_t)$

We obtain space lower bounds for a $(t-1)$-pass streaming algorithm using a t-player game. In that case, space required is at least (communication lower bound)$/(t-1)$.

Disjointness Problem (t-player) Let $x_1, \cdots, x_t \in \{0,1\}^n$, and let x_i be indicator vectors for sets $A_i \subset [n]$. And, define f as:

$$f(x_1, x_2, \cdots x_n) = \begin{cases} 1 & \text{if } \forall i \neq j, A_i \cap A_j \neq \emptyset \\ 0 & \text{if } \forall i \neq j A_i \cap A_j = \{k\} \text{ for some fixed } k \in [n] \\ \text{undefined} & \text{otherwise} \end{cases}$$

We are given the guarantee that f is not undefined, hence it evaluates to 0 or 1. The problem is to decide whether f is 0 or 1.

We now state a theorem which would require about 3 lectures to prove, and is hence left out of the course. The proof also uses an information theoretic approach, known as *information complexity* [4]. The idea is the following chain of inequalities, where Π is the optimal δ-error communication protocol for some function f: $R^{pub}_{\delta}(f) = |\Pi| \geq H(\Pi(\mathbf{X})) \geq I(\mathbf{X}; \Pi(X))$, where $\mathbf{X}$ is the set of inputs given to the t players, and $\Pi(\mathbf{X})$ is the transcript of the communication protocol (or the "communication log") when the input is $\mathbf{X}$ (note that it is a random variable since Π uses randomness). Then we define the *information complexity* $IC_{\mu,\delta}(f)$ as the minimum value $I(\mathbf{X}, \Pi(X))$ achievable by any δ-error protocol Π when $\mathbf{X}$ is drawn from distribution μ. Then we have that $R^{pub}_{\delta}(f) \geq IC_{\mu,\delta}(f)$ for all μ. A variant of this approach was used by [2] to obtain lower bounds for t-player disjointness, with improvements in [3]. The sharp bound was shown in [5], with a later work showing how the arguments in [2] could be strengthened to also get the sharp bound [6].

Theorem 10. $R^{pub}_{1/3}$*(t-player Disjointness)* $= \Omega(n/t)$

Claim 11. *2-approximation of* F_p *requires* $\Omega(n^{1-\frac{2}{p}})$ *bits of space.*

Proof. We reduce from t-player Disjointness, where $t = (2n)^{\frac{1}{p}}$. We do the usual thing where the ith player creates a virtual stream that contains j iff $j \in A_i$. If all the A_i are disjoint, then $F_p \leq n$. If they all however intersect at a single point, then F_p is at least $t^p = 2n$. Thus a 2-approximation of F_p can be used to solve the disjointness instance. For a space-s streaming algorithm the total communication is $s(t-1)$, so the space lower bound we obtain is $\frac{n}{t(t-1)} = \Omega(n/t^2) = \Omega(n^{1-\frac{2}{p}})$. □

References

[1] Noga Alon, Yossi Matias, Mario Szegedy. The Space Complexity of Approximating the Frequency Moments. *J. Comput. Syst. Sci.*, 58(1):137–147, 1999.

[2] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, D. Sivakumar. An information statistics approach to data stream and communication complexity. *J. Comput. Syst. Sci.*, 68(4): 702–732, 2004.

[3] Amit Chakrabarti, Subhash Khot, Xiaodong Sun. Near-Optimal Lower Bounds on the Multi-Party Communication Complexity of Set Disjointness. *IEEE Conference on Computational Complexity*, pgs. 107–17, 2003.

[4] Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, Andrew Chi-Chih Yao. Informational Complexity and the Direct Sum Problem for Simultaneous Message Complexity. *FOCS*, pgs. 270–278, 2001.

[5] Andre Gronemeier. Asymptotically Optimal Lower Bounds on the NIH-Multi-Party Information Complexity of the AND-Function and Disjointness. *STACS*, pgs. 505–516, 2009.

[6] T. S. Jayram. Hellinger Strikes Back: A Note on the Multi-party Information Complexity of AND. *APPROX-RANDOM*, pgs. 562–573, 2009.

[7] T. S. Jayram, Ravi Kumar, D. Sivakumar. The One-Way Communication Complexity of Hamming Distance. *Theory of Computing*, 4(1): 129–135, 2008.

[8] Eyal Kushilevitz, Noam Nisan Communication Complexity *Cambridge University Press*, 1997

[9] David P. Woodruff. Optimal space lower bounds for all frequency moments. *SODA*, pgs. 167–175, 2004.