CS 15-859: Algorithms for Big Data Fall 2019

Lecture 10 — 11/7/2019

Prof. David Woodruff *Scribe: Alireza Samadian*

1 Introduction

We are considering the Gap_∞ problem and give a communication lowerbound for it. In order to do that we introduce **Direct Sum** technique which makes us able to divide the main problem in n dimensional space into multiple smaller problems of 1 dimensional and use the information lowerbound on the 1 dimensional to get a lower bound for the n dimensional problem. The following is the definition of the Gap_∞ problem:

Definition. Alice and Bob are given two n dimensional vectors $x \in 0, \ldots, B^n$ and $y \in 0, \ldots, B^n$ (Alice only has x an Bob only has y). They are promised that either $|x-y|_\infty \leq 1$ or $|x-y|_\infty \geq B$ and their goal is finding out which case is true with high probability.

We are assuming that the communication is 1-way, meaning Alice sends a message to Bob and after that Bob needs to output which of the two cases is true. A lower bound of $\Omega(n/B^2)$ has been shown for the communication complexity of this problem [1, 3].

2 One Dimensional Gap_∞ Problem

Before jumping to the proof for lower bound and explaining direct sum, we first introduce the one dimensional version of the problem and introduce some distribution of the inputs in which the naive algorithm will transfer little information with a $\theta(1)$ communication. This will be a warm up solution and it is intended to show that sometimes there are distributions in which the correct algorithm reveals little information about the input.

Bellow is the definition of the one dimensional Gap_∞

Definition. Alice and Bob are given two integers $x \in 0, \ldots, B$ and $y \in 0, \ldots, B$ (Alice only has x an Bob only has y). They are promised that either $|x-y| \leq 1$ or $|x-y| \geq B$ and their goal is finding out which case is true with high probability.

The only way that $|x-y| \geq B$ would be true is either $(x = B, y = 0)$ or $(x = 0, y = B)$. So one correct deterministic algorithm can send $\theta(1)$ bit indicating if Alice is in either of these two cases and Bob will be able to decide if he should output $|x-y| \leq 1$ or $|x-y| \geq B$.

Now lets consider the following distribution λ which is uniform among all the following values of (x, y).

$$\{(0,0),(0,1),(1,1),(1,2),(2,2),(2,3),\ldots,(B-1,B-1),(B-1,B),(B,B)\}$$

Now lets see what information the message that Alice sends reveals about the input.

$$
\begin{aligned}
I(m;x,y) &= H(x,y) - H(x,y|m) \\
&= \log(2B+1) - (H(x,y|x \in \{0,B\})\mathbf{Pr}[x \in \{0,B\}] + H(x,y|x \notin \{0,B\})\mathbf{Pr}[x \notin \{0,B\}]) \\
&= \log(2B+1) - (\log(3)\frac{3}{2B+1} + \log(2B-2)\frac{2B-2}{2B+1}) \\
&= \log(\frac{2B+1}{2B-2})\frac{2B-2}{2B+1} + \log(2B+1)\frac{3}{2B+1} - \log(3)\frac{3}{2B+1} \\
&= \theta(\frac{log(B)}{B})
\end{aligned}
$$

3 Gap_∞ Problem

In this case, the goal is deciding if $|X-Y|_\infty \leq 1$ or $|X-Y|_\infty \geq B$. Note that this is similar to deciding if $\bigvee_i |X_i - Y_i| \geq B$, because even if in one of these dimensions this happens, then it will be true for the infinity norm. We are going to give a lower bound on the amount of information that a correct algorithm needs to send when the input is coming from some distribution. Using this lowerbound, we can prove a lower bound on the communication complexity.

Consider the following distribution for X and Y where $(X_i, Y_i) \sim \lambda$ are independent uniformly random subject to $|(X_i, Y_i)| \leq 1$. Then among all the correct algorithms for Gap_∞ consider the one that minimizes $I(\psi; X, Y)$ where ψ is the transcript of the protocol (message Alice sends and Bob's output). Let $IC(G) = \inf_\psi I(\psi; X, Y)$, we need to make a lower bound on $IC(G)$.

Theorem 1. $IC(G) = \Omega(\frac{n}{B^2})$.

We will prove this theorem at the end of this section.

In order to give this lower bound, we are going to first assume there is an efficient protocol for Gap_∞, then we are going to show a reduction in which Alice and Bob use this protocol to solve their one bit version of the problem. Then we are arguing that if the main protocol is correct, Alice and Bob should be able to embed their single bit in any dimensions of the vectors and still get the correct result with high probability. This means, the information sent on each dimension should be sufficient for them to retrieve their one dimensional answer.

However, unfortunately Alice and Bob cannot fill the rest of the dimensions randomly without shared randomness. That is because λ is not a product distribution. Therefore, we need another solution. We assume they have some shared randomness D that we will define as follow and then we condition everything on that.

Embedding Step: Alice and Bob put their one bit x and y that are from distribution λ into the i^{th} dimension of X and Y. Then they fill the other dimensions of X and Y as following: we assume both of them have access to $D = ((P_1, V_1), (P_2, V_2), \ldots, (P_n, V_n))$ in some share randomness where:

- P_j is uniform on {Alice, Bob}
- V_j is uniform on $\{1, \ldots, B\}$ if P_j =Alice
- V_j is uniform on $\{0, \ldots, B-1\}$ if P_j =Bob

- For all $j \neq i$, if P_j =Alice then $Y_j = V_j$ and X_j is uniform on $\{V_j - 1, V_j\}$; if P_j =Bob then $X_j = V_j$ and Y_j is uniform on $\{V_j + 1, V_j\}$

Note that conditioned on D, X and Y are independent. Otherwise, they are coming from λ^n distribution. This means Alice and Bob can generate X and Y independently using D and their input bits x and y.

Direct Sum: Using chain rule we have:

$$\begin{aligned} I(M; X, Y) &\geq \sum_i I(M; X_i, Y_i | X_1, Y_1, \ldots, X_{i-1}, Y_{i-1}) \\ &= \sum_i I(M; X_i, Y_i) \qquad \text{since dimmensions are independent} \end{aligned}$$

We also know since it is correctly answering the n dimensional problem then it should be correct for the chosen dimension of Alice and Bob; therefore,

$$I(M; X, Y) \geq \sum_i I(M; X_i, Y_i) \geq IC(g)$$

where $IC(g)$ is the smallest amount of information that a correct one dimensional algorithm will transfer for an input coming from λ distribution.

Now since we are conditioning everything on D we will get the following for the instance constructed in the embedding step:

$$I(M; X_i, Y_i | D) \geq IC(g|D)$$

Therefore,

$$I(M; X, Y | D) = \Omega(n) IC(g|D) \tag{1}$$

. Note that this means all we need is showing a lower bound for $IC(g|D) = \inf_\psi I(\psi; x, y|D)$ where x and y are one dimensional bits and ψ ranges over the transcript of all protocols that are correct with probability 2/3 on the one dimensional input.

This technique is called **Direct Sum**. Now we need to prove the following lemma

Lemma 1. $IC(g|D) = \Omega(\frac{1}{B^2})$

if we prove this lemma then using the Direct Sum technique we can prove Theorem 1. The following subsection is devoted to proof of Lemma 1

3.1 Primitive Problem

Before jumping to the proof of Lemma 1 we first prove the following Lemma about the transcript of a correct protocol for the one dimensional instances.

Lemma 2. *Let $\psi_{a,b}$ denote the transcript of a correct protocol for g when the input is $x = a$ and $y = b$. Then we have the following three claims:*

1. *(Correctness)* $h(\psi_{0,0}, \psi_{0,B})^2 = \Omega(1)$
2. *(1-way Protocol)* $\psi_{a,b}(m, out) = p_a(m).q_{b,m}(out)$
3. *(Pythagorean)* $h^2(\psi_{a,b}, \psi_{c,d}) \geq \frac{1}{2}(h^2(\psi_{a,b}, \psi_{a,d}) + h^2(\psi_{c,b}, \psi_{c,d}))$

Proof. The first two claims are trivial to prove. The first claim is correct because if the distribution of $\psi_{0,0}$ and $\psi_{0,B}$ are very similar then the protocol cannot distinguish between them and therefore it will fail with high probability. The second claim is correct because a message is sent from Alice (who only has access to a) to Bob and then Bob will produce the output only based on b and the message.

The proof of the third claim is a bit longer. Based on Hellinger distance definition we have:

$$1 - h^2(\psi_1, \psi_2) = 1 - \frac{1}{2}\left\|\sqrt{\psi_1} - \sqrt{\psi_2}\right\|_2^2 = 1 - \frac{1}{2}(1 + 1 + 2\left\langle\sqrt{\psi_1}, \sqrt{\psi_2}\right\rangle) = \left\langle\sqrt{\psi_1}, \sqrt{\psi_2}\right\rangle \tag{2}$$

Using this equality we have:

$$\frac{1}{2}(1 - h^2(\psi_{a,b}, \psi_{a,d}) + 1 - h^2(\psi_{c,b}, \psi_{c,d})) = \frac{1}{2}(\left\langle\sqrt{\psi_{a,b}}, \sqrt{\psi_{a,d}}\right\rangle + \left\langle\sqrt{\psi_{c,b}}, \sqrt{\psi_{c,d}}\right\rangle)$$

Note that $\sqrt{\psi_{a,b}}$ is a vector that can be indexed by (m, o) (o is the output). Then using the second claim, we have:

$$\begin{aligned}
&\frac{1}{2}(1 - h^2(\psi_{a,b}, \psi_{a,d}) + 1 - h^2(\psi_{c,b}, \psi_{c,d})) \\
=&\frac{1}{2}\sum_{(m,o)}(\sqrt{p_a(m)q_{b,m}(o)p_a(m)q_{d,m}(o)} + \sqrt{p_c(m)q_{b,m}(o)p_c(m)q_{d,m}(o)}) \\
=&\frac{1}{2}\sum_{(m,o)}(p_a(m)\sqrt{q_{b,m}(o)q_{d,m}(o)} + p_c(m)\sqrt{q_{b,m}(o)q_{d,m}(o)}) \\
=&\sum_{(m,o)}\frac{p_a(m) + p_c(m)}{2}\sqrt{q_{b,m}(o)q_{d,m}(o)} \\
\geq&\sum_{(m,o)}\sqrt{p_a(m) + p_c(m)}\sqrt{q_{b,m}(o)q_{d,m}(o)} \\
=&\left\langle\sqrt{\psi_{a,b}}, \sqrt{\psi_{c,d}}\right\rangle \\
=&1 - h^2(\psi_{a,b}, \psi_{c,d})
\end{aligned}$$

The inequality is because the arithmetic mean is greater than geometric mean. Using the above the third claim follows. ■

Now that we proved this lemma we can have the proof for Lemma 1.

proof of Lemma 1: We are trying to lower bound $IC(g|(P, V)) = \inf_\psi I(\psi; x, y|(P, V))$ where . Based on the definition of conditional mutual information we have:

$$I(\psi; x, y|(P, V)) \geq \frac{1}{2}E_v[I(\psi; x, y|(\text{Alice}, v)) + I(\psi; x, y|(\text{Bob}, v))]$$

Consider the case that P =Alice and $V = v$, then there are two possible cases for y $(v-1, v)$ and based on those possibilities we have two different distribution of values for ψ ($\psi_{v-1,v}$ and $\psi_{v,v}$). Similar thing for the case of Bob. From previous lecture we know Jenson-Shannon lower bounds Information, meaning:

$$I(X;B) \geq D_{JS}(X|B=0, X|B=1)$$

where B is a binary random variable. Using this we can get the followings:

$$I(\psi; x, y|(\text{Alice}, v)) \geq D_{JS}(\psi_{v-1,v}, \psi_{v,v})$$

and

$$I(\psi; x, y|(\text{Bob}, v)) \geq D_{JS}(\psi_{v,v}, \psi_{v,v+1})$$

Therefore,

$$I(\psi; x, y|(P, V)) \geq \frac{1}{2} E_v[D_{JS}(\psi_{v-1,v}, \psi_{v,v}) + D_{JS}(\psi_{v,v}, \psi_{v,v+1})]$$

We also know from the previous lectures that $D_{JS}(\psi_{v,v}, \psi_{v,v+1}) \geq h^2(\psi_{v,v}, \psi_{v,v+1})$; consequently,

$$\begin{aligned}
&I(\psi; x, y|(P, V)) \\
&\geq \frac{1}{2} E_v[h^2(\psi_{v-1,v}, \psi_{v,v}) + h^2(\psi_{v,v}, \psi_{v,v+1})] \\
&\geq \frac{1}{2B} \sum_v \left(\left\| \sqrt{\psi_{v-1,v}} - \sqrt{\psi_{v,v}} \right\|^2 + \left\| \sqrt{\psi_{v,v}} - \sqrt{\psi_{v,v+1}} \right\|^2 \right) \\
&\geq \frac{1}{2B^2} \left(\sum_v \left\| \sqrt{\psi_{v-1,v}} - \sqrt{\psi_{v,v}} \right\| + \left\| \sqrt{\psi_{v,v}} - \sqrt{\psi_{v,v+1}} \right\| \right)^2 && \text{(cauchy-schwarz)} \\
&\geq \frac{1}{2B^2} \left(\sum_v \left\| \sqrt{\psi_{v,v}} - \sqrt{\psi_{v+1,v+1}} \right\| \right)^2 && \text{(triangle inequality)} \\
&\geq \frac{1}{2B^2} \left(\left\| \sqrt{\psi_{0,0}} - \sqrt{\psi_{B,B}} \right\| \right)^2 && \text{(triangle inequality)} \\
&\geq \frac{1}{4B^2} \left(\left\| \sqrt{\psi_{0,0}} - \sqrt{\psi_{0,B}} \right\|^2 + \left\| \sqrt{\psi_{B,0}} - \sqrt{\psi_{B,B}} \right\|^2 \right) && \text{(Pythagorean Lemma 2)} \\
&\geq \frac{1}{4B^2} (\Omega(1) + \Omega(1)) && \text{(Correctness Lemma 2)} \\
&= \Omega(\frac{1}{B^2})
\end{aligned}$$

■

3.2 Proof of Theorem 1

Using Lemma 1 we know $IC(g|D) = \Omega(\frac{1}{B^2})$; we can use this lower bound and combining with the embedding step and direct sum technique we know:

$$IC(G|D) = \inf_{\psi} I(\psi; X, Y|D) \geq \inf_{\psi} \sum_i I(\psi; X_i, Y_i|D) \geq n \inf_{\psi} I(\psi; X_i, Y_i|D) = \Omega(\frac{n}{B^2})$$

■

Corollary 1. *The communication complexity of Gap_∞ is $\Omega(\frac{n}{B^2})$*

Proof. This is because by having shared randomness the communication complexity does not increase and using shared randomness we have shown $IC(G|D) = \Omega(\frac{n}{B^2})$. Then we know this would be a lower bound for the message size. This is because of the result from the previous lecture:

$$|M| \geq H(M) \geq H(M|D) \geq H(M|D) - H(M|X, Y, D) = I(M; X, Y|D)$$

■

Using similar technique we can get a $\Omega(n)$ lower bound for **Set Disjointness** Problem [1].

Remark 1. [2] has shown sometimes we can get a better lower bound that what we get using direct sum. This is because direct sum solves the smaller problems using a constant probability. There are some problems in which if we want a constant probability for larger instance, we need $1 - \frac{1}{n}$ probability of success for the smaller instances. The example is 1-way communication complexity of Equality.

Remark 2. Direct sum is a nice technique but in many cases the problem cannot be split into simpler smaller problems. For instance there is no known embedding step in Gap-Hamming problem.

References

[1] Ziv Bar-Yossef, Thathachar S Jayram, Ravi Kumar, and D Sivakumar. An information statistics approach to data stream and communication complexity. *Journal of Computer and System Sciences*, 68(4):702–732, 2004.

[2] Marco Molinaro, David P Woodruff, and Grigory Yaroslavtsev. Beating the direct sum theorem in communication complexity with implications for sketching. In *Proceedings of the twenty-fourth annual ACM-SIAM symposium on Discrete algorithms*, pages 1738–1756. SIAM, 2013.

[3] Michael Saks and Xiaodong Sun. Space lower bounds for distance approximation in the data stream model. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 360–369. ACM, 2002.