Communication Complexity

Lecturer: Alexander Lindenbaum and Yunya Zhao

1 Review and Preliminaries

- Entropy of a random variable X

$$\mathsf{H}(X) = \sum_x p(x) \cdot \log \frac{1}{p(x)} = \mathop{\mathbb{E}}_{p(x)} \left[\log \frac{1}{p(x)} \right]$$

- Conditional Entropy

$$\mathsf{H}(Y|X) = \mathop{\mathbb{E}}_{p(xy)} \left[\log \frac{1}{p(y|x)} \right] = \mathop{\mathbb{E}}_{p(x)} [\mathsf{H}(Y|X=x)]$$

- Chain Rule of Entropy

$$\mathsf{H}(XY) = \mathsf{H}(X) + \mathsf{H}(Y|X)$$

- Mutual Information

$$\mathsf{I}(A;B) = \mathsf{H}(A) - \mathsf{H}(A|B)$$

- Conditional Mutual Information

$$\mathsf{I}(A;B|C) = \mathsf{H}(A|C) - \mathsf{H}(A|BC)$$

- Chain Rule of Mutual Information

$$\mathsf{I}(AB;C) = \mathsf{I}(A;C) + \mathsf{I}(B;C|A)$$

- Chain Rule of Conditional Mutual Information

$$\mathsf{I}(AB;C|D) = \mathsf{I}(A;C|D) + \mathsf{I}(B;C|AD)$$

2 Information Complexity

Analogous to Communication Cost and Communication Complexity: **Information Cost** is related to the amount of information gained through the execution of a communication protocol π, **Information Complexity** is related to a function f (a problem) over all protocols that computes it.

Definition 1 (Transcript of a protocol)**.** *Given a protocol π, the **transcript** $\boldsymbol{\pi}(\mathbf{X}, \mathbf{Y})$ is the concatenation of the public randomness with all the messages that are sent during the execution of π on input X, Y*

Definition 2 (Internal information cost)**.** *Internal information cost* $\mathsf{IC}^i_\mu(\pi)$ *is how much each party learns about the other party's input during the execution of* π

$$\mathsf{IC}^i_\mu(\pi) = I(X;\pi(X,Y)|Y) + I(Y;\pi(X,Y)|X)$$

Definition 3 (External information cost)**.** *External information cost* $\mathsf{IC}^{ext}_\mu(\pi)$ *is how much information an outside observer learns about both parties' input just by looking at Alice and Bob chat*

$$\mathsf{IC}^{ext}_\mu(\pi) = I(XY;\pi(X,Y))$$

For any protocol, the internal information cost is no larger than the external information cost:

Theorem 1. *For protocol* π *and distribution* μ*, we have*

$$\mathsf{IC}^i_\mu(\pi) \le \mathsf{IC}^{ext}_\mu(\pi)$$

Proof. Let ω be any fixed prefix of the transcript of length $i-1$. An observer always learns something new at each round, and that amount is

$$I(XY;\pi(X,Y)_i|\pi(X,Y)_{\le i-1} = \omega)$$

$$= I(X;\pi(X,Y)_i|\pi(X,Y)_{\le i-1} = \omega) + I(Y;\pi(X,Y)_i|X\pi(X,Y)_{\le i-1} = \omega)$$
$$\ge I(X;\pi(X,Y)_i|Y\pi(X,Y)_{\le i-1} = \omega) + I(Y;\pi(X,Y)_i|X\pi(X,Y)_{\le i-1} = \omega)$$

NOTE: if μ is a product distribution, $\mathsf{IC}^i_\mu(\pi) = \mathsf{IC}^{ext}_\mu(\pi)$ □

3 Preliminaries

Given a function $f(x,y)$ and a distribution μ on inputs to f

- The communication complexity $D^\mu_\rho(f)$, maximum number of bits communicated by a protocol that computes f with error ρ
- $D^{\mu,n}_\rho(f)$, the communication involved in the best protocol that computes f on n **independent** pairs of input (x,y) drawn from μ, and getting the answer correct except an error ρ **on each coordinate**.
- Note that the above is different from $D^{\mu^n}_\rho(f^n)$, and

$$D^{\mu,n}_\rho(f) \le D^{\mu^n}_\rho(f^n)$$

4 Direct Sum

The **direct sum question** is about the complexity of solving *several* copies of a given problem. In communication complexity, it can be phrased as follows: given function

$$f : \{0,1\}^m \times \{0,1\}^m \longrightarrow \{0,1\}$$

define

$$f^n : (\{0,1\}^m)^n \times (\{0,1\}^m)^n \longrightarrow \{0,1\}^n$$

to be

$$f^n((x_1, ..., x_n), (y_1, ..., y_n)) = (f(x_1, y_1), ..., f(x_n, y_n))$$

What is the relationship between the communication costs of f and f^n.

The direct sum problem is important: direct sum theorems, together with a lower bound on the (easier-to-reason-about) "primitive" problem, yield a lower bound on the composite problem in a "black-box" fashion (a method also known as *hardness amplification*) [6]

- Ex. Karchmer-Raz-Wigderson: $\mathsf{P} \neq \mathsf{NC}^1$ if circuit depth has strong direct sum (there are inherently sequential problems)

However, direct sum theorems are very sensitive to models, and that is why we don't much about direct sum problems for many models.

Trivially, the communication complexity for f^n is most n times the communication complexity of f.

$$D(f^n) \leq n \cdot D(f)$$

Strong Direct Sum Conjecture "the naive is the optimal"

$$D_\rho^{\mu^n}(f^n) \overset{?}{=} \Omega(n) \cdot D_\rho^\mu(f)$$

One direction is trivial, need to prove the other direction

Direct Sum Theorem for Simultaneous Communication (the equality function) [3]

The equality problem of n-bit string have SM complexity $\Theta(\sqrt{n})$

$$D(\mathsf{EQ}_n^m) = \Omega(m\sqrt{n})$$

5 Why use information theoretic tools

- [3] used information theoretic tools to arrive at direct sum because:
- Information Complexity has a nice direct sum property

$$IC^n(f) \geq n \cdot IC(f)$$

- The above property bridges together direct sum of communication:

$$D^n(f) \geq IC^n(f) \geq n \cdot IC(f) \ \ ??? \ \ n \cdot D(f)$$

6 Information Equals Amortized Communication [2]

The internal information cost (namely the information revealed to the parties) involved in computing any relation or function using a two party interactive protocol is exactly equal to the amortized communication complexity of computing independent copies of the same relation or function. If a function's information cost is smaller than its communication complexity, then multiple copies of the function can be computed **more efficiently in parallel than sequentially**.

- The **amortized communication complexity**

$$\lim_{n\to\infty} \frac{D_\rho^{\mu,n}(f)}{n}$$

- Information (internal) equals amortized communication complexity:

$$\lim_{n\to\infty} \frac{D_\rho^{\mu,n}(f)}{n} = IC_\mu(f)$$

7 Information Complexity Direct Sum Theorem [1]

Theorem 2. *For every μ, f, n, let π be a protocol realizing $D_\rho^{\mu,n}(f)$. Then there exists a protocol τ computing f with error ρ on inputs drawn from μ such that $CC(\tau) = CC(\pi)$, and $IC_\mu^i(\tau) \leq \frac{IC^i_{\mu^n}(\pi)}{n} \leq \frac{D_\rho^{\mu,n}(f)}{n}$ $(\leq \frac{D_\rho^{\mu^n}(f^n)}{n})$*

Proof. First let us assume that π only uses private randomness (can easily extend to cover public randomness case). The new protocol $\tau(x, y)$ is defined as follows:

- the parties publicly sample J uniformly at random from [n].
  J is understood as an index.
- The parties publicly sample $X_1, ..., X_{J-1}$ and $Y_{J+1}, ..., Y_n$.
- The first party privately samples $X_{J+1}, ..., X_n$ conditioned on the corresponding Y's; The second party does similar.
- The parties run the old protocol π on $X_1, ..., X_n, Y_1, ..., Y_n$ and output the result computed for the J'th coordinate. (i.e. viewing $X_J = x$, $Y_J = y$)

Analyze the protocol: observe CC and bounded error: $CC(\pi) = CC(\tau)$, and error is bounded by ρ. It remains to bound $IC_\mu^i(\tau) = I(X; \tau|Y) + I(Y; \tau|X)$. NOTE: X, Y are r.v. for τ's inputs (sampled according to μ).

Let's bound the first term:

$$\begin{aligned}
I(X : \tau|Y) &\leq I(X : \tau Y_1 \cdots Y_n|Y) \\
&= I(X; JX_1 \cdots X_{J-1}Y_1 \cdots Y_n\pi|Y) \\
&= I(X; JX_1 \cdots X_{J-1}Y_1 \cdots Y_n|Y) + I(X_J; \pi|JX_1 \cdots X_{J-1}Y_1 \cdots Y_n) \\
&= I(X_J; \pi|JX_1 \cdots X_{J-1}Y_1 \cdots Y_n)
\end{aligned}$$

Expanding the expectation according to J, apply Chain Rule:

$$\begin{aligned} I(X;\tau|Y) &\leq (1/n)\sum_{j=1}^{n} I(X_j;\pi|X_1\cdots X_{j-1}Y_1\cdots Y_n) \\ &= I(X_1\cdots X_n;\pi|Y_1\cdots Y_n)/n \end{aligned}$$

Similarly we can bound $I(Y;\tau|X) \leq I(Y_1\cdots Y_n;\pi|X_1\cdots X_n)/n$, and thus $\mathsf{IC}^i_\mu(\tau) \leq \mathsf{IC}^i_{\mu^n}(\pi)/n \leq \mathsf{CC}(\pi)/\mathsf{n}$. □

Now we have a way of taking a protocol π for n copies of f and constructing a protocol τ for a single copy. But the communication complexity of τ is equal to that of π, even though we are only computing one copy of f. This motivates a method to compress protocols with high communication but low (internal) information complexity.

Theorem 3. *For every distribution μ, every protocol π, and every $\epsilon > 0$, there exists functions π_x, π_y, and a protocol τ such that $|\pi_x(X,\tau(X,Y)) - \pi(X,Y)| < \epsilon$, $\Pr[\pi_x(X,\tau(X,Y)) \neq \pi_y(Y,\tau(X,Y))] < \epsilon$ and*

$$CC(\tau) \leq O\left(\sqrt{CC(\pi)\cdot IC^i_\mu(\pi)}\frac{\log(CC(\pi)/\epsilon}{\epsilon}\right).$$

Proof. In τ, Alice and Bob will independently guess π's transcript. Since both parties will likely be incorrect (Alice does not know y, Bob does not know x), they will communicate with as few bits as possible to correct their guesses. More specifically, let $M = m_1m_2\cdots m_{CC(\pi)}$ denote the random variable for π's transcript (wlog we assume that π communicates $CC(\pi)$ bits). Let $m_{<i}$ denote the first $i-1$) bits sent of M. Define

$$\gamma(m_{<i}) = \Pr[M_i = 1|X = x, Y = y, M_{<i} = m_{<i}].$$

This defines a distribution over M. One can sample from this distribution in the following way:

- Obtain $\rho_1,\ldots,\rho_{CC(\pi)} \sim [0,1]$ uniformly and identically by reading random bits from a public source.
- Set $m_1 = 1$ if $\rho_1 < \gamma(m_{<1}) = \Pr[M_1 = 1|X = x, Y = y)$. Otherwise $m_1 = 0$.
- Set $m_i = 1$ if $\rho_i < \gamma(m_{<i})$ for $i = 2,\ldots,CC(\pi)$ in that order.

Indeed if Alice and Bob could sample from this distribution, then they would each have an exact copy M which is distributed exactly as π's transcript. Then they could simply run the protocol π internally and output $\pi(x,y)$. The problem remains that Alice does not know y, Bob does not know x.

Instead, Alice samples from the distribution

$$\gamma^A(m_{<i}) = \Pr[M_i = 1|X = x, M_{<i} = m_{<i}].$$

Likewise Bob samples from

$$\gamma^B(m_{<i}) = \Pr[M_i = 1|Y = y, M_{<i} = m_{<i}].$$

Generally, these distribution are not equal to the true distribution. But say Alice has guessed $m_{<i}$ which is distributed according to π, and it is Alice's turn to speak on the ith bit. Then $\gamma^A(m_{<i}) = \gamma(m_{<i})$. This is because Alice has all the information that she would have in π right before sending the ith bit: the correctly sampled transcript up to i and x. Likewise if it is Bob's turn to speak and he has $m_{<i}$ then $\gamma^B(m_{<i}) = \gamma(m_{<i})$.

So the protocol τ is as follows:

- Alice and Bob produce $m_1^A \cdots m_{CC(\pi)}^A$ and $m_1^B \cdots m_{CC(\pi)}^B$, respectively according to γ^A, γ^B.
- Using binary search and hashing, Alice and Bob communicate to find the first index i where $m_i^A \neq m_i^B$.
- If it is Alice's turn to speak on the ith bit (in π), then Bob flips m_i^B and recomputes from $i+1$ to $CC(\pi)$. Otherwise Alice recomputes.
- Repeat until $m^A = m_B$. Then Alice or Bob compute $\pi(x, y)$ from the transcript and output.

It takes $O(\log(CC(\pi)/\epsilon))$ bits of communication find the first i where $m_i^A \neq m_i^B$, with error $\epsilon/2$. Now we bound the total number of mistakes to correct, in probability.

A mistake occurs at the ith bit if ρ_i falls in between $\gamma^A(m_{<i})$ and $\gamma^B(m_{<i})$. Suppose Alice is sending the ith bit in π. Then the probability of a mistake occuring at i, given $m_{<i}$, is

$$\begin{aligned}
&\mathbb{E}\Big[\big|\gamma^A(m_{<i}) - \gamma^B(m_{<i})\big|\Big] \\
\leq &\mathbb{E}\Big[\big|\Pr[M_i = 1|X = x, M_{<i} = m_{<i}] - Pr[M_i = 1|Y = y, M_{<i} = m_{<i}]\big|\Big] \\
\leq &\sqrt{I(M_i : X|YM_{<i})} = \sqrt{I(X : M_i|YM_{<i})}
\end{aligned}$$

By the fact that $\mathbb{E}_b[|p(a|b) - p(a)|] \leq \sqrt{I(A : B)}$ (follows from Pinsker's inequality). Similarly if Bob sends the ith bit then the probability of a mistake at i is bounded by $\sqrt{I(Y : M_i|XM_{<i})}$. Then a crude upper bound on the expected number of mistakes made is

$$\begin{aligned}
&\sum_{i=1}^{CC(\pi)} \sqrt{I(X : M_i|YM_{<i}) + I(Y : M_i|XM_{<i})} \\
&\leq \sqrt{CC(\pi)} \cdot \sqrt{\sum_{i=1}^{C} I(X : M_i|YM_{<i}) + I(Y : M_i|XM_{<i})}
\end{aligned}$$

by Cauchy-Shwarz;

$$= \sqrt{CC(\pi)} \cdot \sqrt{I(X : M|Y) + I(Y : M|X)} = \sqrt{IC_\mu^i(\pi)CC(\pi)}.$$

Finally by a simple application of Markov's inequality, the probability that the total number of bits communicated is greater than $2/\epsilon \cdot O(\sqrt{IC_\mu^i(\pi)CC(\pi)} \cdot \log(CC(\pi)/\epsilon))$ is at most $\epsilon/2$. □

Now we have enough to prove a direct sum theorem.

Theorem 4. *For every $\epsilon > 0$,*

$$D_\rho(f^n) \cdot \log(D_\rho(f^n)/\epsilon) \geq \Omega(R_{\rho+\epsilon}(f) \cdot \epsilon\sqrt{n}.$$

Proof. Let π be any protocol computing f^n with inputs drawn from μ^n with error probability $\leq \rho$. By Theorem 2, we have a protocol τ computing f with error ρ such that $CC(\tau) = CC(\pi)$ and $IC_\mu^i(\tau) \leq IC_\mu^i(\pi)/n$. Using Theorem 3, we compress τ to get τ', a protocol computing f with error $\rho + \epsilon$ and communication

$$CC(\tau') \leq O\left(\frac{CC(\pi)\log(CC(\pi)/\epsilon)}{\epsilon\sqrt{n}}\right).$$

We have that $CC(\tau') \geq R_{\rho+\epsilon}(f)$, so for all π computing f^n,

$$CC(\pi)\log(CC(\pi)/\epsilon) \geq \Omega(R_{\rho+\epsilon}(f) \cdot \epsilon\sqrt{n}.$$

□

Corollary: $CC(T^n) = \tilde{\Omega}(\sqrt{n} \cdot CC(T))$.

Proof. By applying Yao's min-max principle. □

8 Tight Direct Sum Bounds and Separations

The question remains: can we show that $CC(T^n) = \Theta(n \cdot CC(T))$? The answer, unfortunately, is *no*. This is because we know of classes of problems (and corresponding input distributions) with communication complexity of order exponential the information complexity. If $IC(T) = k$ and $CC(T) = 2^{\Omega(k)}$, then the amortized complexity tells us that we can never have $CC(T^n) = \Omega(n \cdot CC(T))$.

- In 2014, Ganor, Kol, and Raz [4] showed that the *bursting noise game*, with fixed distribution, has information complexity k and communication complexity $2^{\Omega(k)}$.
- In 2018, Rao and Sinha [5] give a simpler problem, the k-ary pointer jumping problem, which also has an exponential difference in information complexity and communication complexity.

References

[1] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. "How to compress interactive communication." *STOC* 2010. 4

[2] Mark Braverman and Anup Rao. "Information Equals Amortized Communication." *FOCS* 2011. 4

[3] Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew Yao. "Information Complexity and the Direct Sum Problem for Simultaneous Message Complexity." *FOCS* 2001. 3

[4] Anat Ganor, Gillat Kol, Ran Raz. ”Exponential Separation of Information and Communication.” *FOCS* 2014. 7

[5] Anup Rao, Makrand Sinha. ”Simplified Separation of Information and Communication.” *Theory of Computing* 2018. 7

[6] Omri Weinstein. ”Information Complexity and the Quest for Interactive Compression (A Survey).” *FOCS* 2015.

3