CAS CS 591 B: Communication Complexity

Prof. Mark Bun — Fall 2019

Lecture Notes 11:

Direct Sum

Reading.

- Rao-Yehudayoff Chapter 7

Does solving k instances of a given computational task require more resources than are required to solve a single task? Direct sum theorems (as well as related direct product theorems and XOR lemmas) address this question. The direct sum question is one of the most basic questions one can ask in any computational model and here we'll see some answers to it in communication and information complexity.

For a communication problem $f : X \times Y \to \{0,1\}$, let $f^k : X^k \times Y^k \to \{0,1\}^k$ be the function $f^k(x_1, \dots, x_k, y_1, \dots, y_k) = (f(x_1, y_1), \dots, f(x_k, y_k))$. A randomized protocol computes f^k to error ε if for all $(x, y) \in X^k \times Y^k$ we have $\Pr[\Pi(x, y) = f^k(x, y)] \geq 1 - \varepsilon$. Note that if we have a protocol computing f to error ε with communication cost c, then we can compute f^k to error $(1-\varepsilon)^k$ with communication cost $k \cdot c$.

- Note that naïve repetition of a protocol for f both increases the communication cost by a factor of k and decreases the success probability exponentially in k. (Strong) direct product theorems identify situations where both an increase in communication and exponential decrease in success probability are necessary. For instance, a direct product theorem might show that if c is the communication cost of computing f to error $1/3$, then computing f^k to error $2^{-\Omega(k)}$ requires communication $\Omega(k \cdot c)$.

- Related to direct product theorems are XOR lemmas which address the task of computing the boolean function $\mathrm{XOR}(f(x_1, y_1), \dots, f(x_k, y_k))$. Intuitively, successfully computing the XOR of k copies of f should necessitate computing every copy simultaneously. An XOR lemma might say that computing $\mathrm{XOR}_k \circ f$ to error $1/2 + 2^{-\Omega(k)}$ requires communication $\Omega(k \cdot c)$.

- Direct sum theorems are weaker than direct product theorems, and ask whether computing f^k requires more resources than computing f itself for some fixed success probability ε. A direct sum theorem may say that computing f^k to error $1/3$ requires communication $\Omega(k \cdot c)$.

1 Additivity of Information Cost

Information cost behaves elegantly with respect to k-fold composition.

Theorem 1. *Let f be any communication problem and let μ be a distribution over the domain of f. Then for every $\varepsilon > 0$,*

$$IC^{\varepsilon}_{\mu,k}(f^k) = k \cdot IC^{\varepsilon}_{\mu}(f).$$

Here, the notation $IC^{\varepsilon}_{\mu,k}(f^k)$ refers to the least information cost of a protocol computing f^k which has success probability $1-\varepsilon$ on each individual copy.

Proof. Fix $\varepsilon > 0$. The direction $IC_{\mu,k}(f^k) \leq k \cdot IC_\mu(f)$ is obvious since we can just run a protocol Π computing f on each copy independently. For the other direction, let Π be a protocol computing f^k. We will use it to construct a protocol Π' computing f with information cost $IC_{\mu,k}(\Pi)/k$. The protocol is as follows. On inputs (x, y), Alice and Bob use public randomness to sample an index $i^* \in [n]$ and set $a_{i^*} = x$ and $b_{i^*} = y$. They also publicly sample $a_1, \ldots, a_{i^*-1} \sim \mu_x$ and $b_{i^*+1}, \ldots, b_n \sim \mu_y$ i.i.d. Then for $i > i^*$, Alice privately samples $a_i \sim \mu_x | y = b_i$ and for $i < i^*$, Bob privately samples $b_i \sim \mu_y | y = a_i$. The parties then execute the protocol $\Pi(a, b)$ and output the value computed in the i^*-th coordinate.

We now argue that Π' has low internal information cost. We calculate

$$\begin{aligned}
I(y; \Pi' | x) &\leq I(y; \Pi', a | x) \\
&= I(y; \Pi, i^*, a, b_{i^*+1}, \ldots, b_n | x) \\
&= I(y; i^*, a, b_{i^*+1}, \ldots, b_n | x) + I(y; \Pi | i^*, a, b_{i^*+1}, \ldots, b_n) \\
&= I(y; \Pi | i^*, a, b_{i^*+1}, \ldots, b_n) \\
&= \frac{1}{k} \sum_{i=1}^{k} I(b_{i^*}; \Pi | a, b_{i+1}, \ldots, b_n, i = i^*) \\
&= \frac{1}{k} \sum_{i=1}^{k} I(b_i; \Pi | a, b_{i+1}, \ldots, b_n) \\
&= \frac{1}{k} I(b; \Pi | a)
\end{aligned}$$

by the chain rule. □

2 Direct Sum for Randomized Communication

Theorem 2. *If* $\mathbf{BPP^{cc}}(f) = c$, *then* $\mathbf{BPP^{cc}}(f^k) \geq \Omega(c\sqrt{k}/\log c)$.

To prove this theorem, we need a different compression result due to Braverman, Barak, Chen, and Rao:

Theorem 3. *A protocol with communication cost c and information cost I can be compressed to a protocol with communication cost*

$$O(\sqrt{Ic}\log(c/\varepsilon)/\varepsilon)$$

with error ε.

Proof of Theorem 2. By Yao's principle, there is a distribution μ such that $D^{1/3}_{\mu}(f) \geq c$. Suppose Π is a protocol computing f^k with success probability $3/4$ and length ℓ. Then there exists a fixing of the randomness r in Π such that Π_r computes f^k with success probability $3/4$ over inputs drawn from $\mu^{\otimes k}$. Now let Π' be the protocol used in the proof of Theorem 1. Then Π' computes f over μ with success probability at least $3/4$ and has internal information cost

$$IC_\mu(\Pi') = \frac{1}{k} IC_{\mu^{\otimes k}}(\Pi_r) \leq \frac{\ell}{k}.$$

Compressing Π' according to Theorem 5 gives a protocol Π'' with communication

$$O\left(\sqrt{\frac{\ell}{k}\cdot\ell}\cdot\log\ell\right) = O(\ell\log\ell/\sqrt{k})$$

and error less than $1/3$. Hence $\ell\log\ell/\sqrt{k} \geq \Omega(c)$, so $\ell \geq \Omega(c\sqrt{k}/\log c)$. □

3 Information Equals Amortized Communication

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

$$IC^{\varepsilon}_{\mu}(f) = \lim_{k\to\infty}\frac{1}{k}\cdot D^{\varepsilon}_{\mu,k}(f^k)$$

where $D^{\varepsilon}_{\mu,k}(f^k)$ denotes the least cost of a deterministic protocol computing f^k over $\mu^{\otimes k}$ with success probability at least $1-\varepsilon$ on every copy.

To prove this theorem, we recall Braverman and Rao's compression for interactive protocols:

Theorem 5. *Let Π be a private coin protocol with r rounds and internal information cost I. Then there is a public coin protocol τ such that with probability at least $1-\varepsilon$,*

- *$\tau(x,y)$ (can be used to reconstruct a transcript which) has the same distribution as $\Pi(x,y)$ for every x, y and*
- *The expected communication of τ is $I + O(r\log(r/\varepsilon) + \sqrt{Ir})$.*

Proof sketch. From Theorem 1 we have

$$\frac{1}{k}\cdot D^{\varepsilon}_{\mu,k}(f^k) \geq \frac{1}{k}\cdot IC^{\varepsilon}_{\mu^{\otimes k}}(f^k) = IC^{\varepsilon}_{\mu}(f).$$

We now prove the other direction. Let $\delta > 0$ and let Π be a protocol computing f to error $\varepsilon - \delta$ in r rounds. Let Π^k denote this protocol applied to k copies of f independently. Note that the number of rounds of Π^k is still r (since we can batch calls to Π together) and that

$$IC_{\mu^{\otimes k}}(\Pi^k) = k\cdot IC_{\mu}(\Pi).$$

Now we compress Π^k to another protocol Π' which uses expected communication

$$k\cdot IC_{\mu}(\Pi) + O(r\log(r/\delta)) + \sqrt{k\cdot r\cdot IC_{\mu}(\Pi)}$$

and additional error δ. Moreover, the communication cost of Π' concentrates around its expectation, so we can obtain a similar bound on its worst case communication. Sending $k\to\infty$ and $\delta\to 0$ gives the result. □