| CS 229r Information Theory in Computer Science | March 28, 2019 |
|---|---|
| Lecture 16 | |
| *Instructor: Madhu Sudan* | *Scribe: Kevin Rao* |

# 1 Today's Agenda

## 1.1 Topics

- Conclude: Compressed Interactions
- Start: Information-Amortized Communication

## 1.2 Logistics

Project presentations likely May 1

# 2 Review

**Theorem 1.** *If some protocol $\pi$ has $CC(\pi) \leq k$ and $IC(\pi) \leq I$ then it can be simulated by another protocol $\pi'$ with $CC(\pi') \leq O(\sqrt{kI})$*

## 2.1 Correlated Sampling

Alice has as input a distribution $P$ and Bob has a distribution $Q$ both supported on $\Omega$ and some shared randomness $R$. Without any communication, they output $\omega_A \sim P$ and $\omega_B \sim Q$ and we wish to minimize $\Pr[\omega_A \neq \omega_B]$

**Exercise 2.** *Suppose Alice and Bob both receive $(P, Q)$, but still must output according to $P$ and $Q$ respectively. Show that* $\min \Pr[\omega_A \neq \omega_B] = \delta(P, Q)$.

Last time we showed that we can achieve $\Pr[\omega_A \neq \omega_B] \leq 2\delta(P, Q)$ and that this bound is tight up to a factor of 2, as Alice only knows $P$ and Bob only knows $Q$. If Alice and Bob do not both know $P, Q$, the factor of 2 matters.

# 3 Simulating a Protocol

Today we wish to come up with a way to simulate a general interaction. We look at the protocol tree, where for each step of the interaction Alice and Bob follow some instructions to go left and right with some probability on the nodes they control.

Goal: Alice and Bob both output a leaf that looks like it was produced with all correct steps

The following three distributions matter:

- If node $u$ is at level $i$, $P_u = \pi_i | \pi_{<i}, x, y$. This quantity is only known to one player at any given node (whichever player "owns" that node)
- $P_u^A = \pi_i | \pi_{<i}, x$
- $P_u^B = \pi_i | \pi_{<i}, y$

We have the following two tensions:

- $P_u^A$ and $P_u^B$ are very close for all $u$. This is good news for us, because we don't need to communicate very much and still probably output correctly.
- $P_u^A$ and $P_u^B$ are far apart. This means each action you take (left or right) leaks information about the inputs and path, so $IC(\pi)$ is high.

This is our general understanding: either distributions are close or far at certain nodes.

## 3.1 BBCR Simulation Protocol

The simulating protocol $\pi'$ put forth by [BBCR] is as follows:

1. $\pi'$ starts at the root, and we want to sample a leaf according to distribution $P_u$ (we make all decisions simultaneously). To do this, on all nodes Alice and Bob both choose a left or right (even if they don't own the node) according to $P_u^A$ and $P_u^B$ with correlated sampling. If the two disagree on a move with high probability, it is because the distributions are far.

   Note: we are making $2^k$ correlated choices for a depth $k$ graph. This is extremely inefficient, but we haven't communicated yet, and that is what we ware trying to minimize so it's fine.

2. - Pick the path from root to leaf for Alice $\ell_A$
   - Pick the path from root to leaf for Bob $\ell_B$.

   If Alice and Bob have the same node, they're in business. If the nodes are different, we must backtrack and binary search for where Alice and Bob began to disagree.

   Thus, the next step is

3. Alice and Bob say, "This is where I think we should end", and each encode their leaf with some k bit representation. They now run an equality testing protocol on their leaves.

4. Binary search for the last node that they share (least common ancestor) by looking first at level $\frac{k}{2}$ and so on. With some work, we see that this requires $O(\log k)$ communication, as we have $k$ choices for where we diverge and binary search takes $\log k$ communication.

5. Call the last common ancestor $u_1$. Now we recursively compute the new root $\pi'(u_1)$.

**Exercise 3.** *Show that the least common ancestor computation will indeed require* $\log k$ *bits of communication.*

Having done this, we claim that we have correctly picked a path. Hopefully, each step of the process takes us fairly far down the tree. If so, the protocol is good. Otherwise, we can simply say there is high information cost.

### 3.1.1 Analysis

We claim that the number of iterations (times we start from a root and do correlated sampling) is at most $\sqrt{kI}$. We can assume that we make at least one step at each iteration, since if Alice and Bob disagree they defer to the person who actually "owns" that step.

*Proof.* We define disagreements to be nodes on the relevant path where Alice and Bob pick different directions. Say $Z_i$ is some indicator r.v. which is 1 if in our process we saw a disagreement at level $i$. The expected number of iterations is $\mathbb{E}\left[\sum_{i=0}^{k} Z_i\right]$

Note: if I tell you some $u$ is on the path, then Pr[disagreement at u|passing through u] $\leq 2\delta(P_u^A, P_u^B)$ (this follows from the correlated sampling procedure).

Thus, $\mathbb{E}\left[\sum_{i=0}^{k} Z_i\right] \leq 2\sum_i \mathbb{E}_u\left[\delta(P_u^A, P_u^B)\right]$, where $u$ is the chosen node at level $i$.

How does this relate to information cost?

Recall that $IC(\pi) = I(\pi; x|y) + I(\pi; y|x) = \sum_i I(\pi_i; x|y, \pi_{<i}) + I(\pi_i; y|x, \pi_{<i})$. Define $v_i \triangleq I(\pi_i; x|y, \pi_{<i}) + I(\pi_i; y|x, \pi_{<i})$. We compare $v_i$ to $\mathbb{E}_u[\delta(P_u^q, P_u^B)]$ and see that $v_i$ roughly measures expected divergence over $u$, so we are left with $\mathbb{E}_u[D(P_u^A||P_u^B)]$ vs $\mathbb{E}[\delta(P_u^A, P_u^B)]$. We want to show that the latter is an upper boud of the former.

**Theorem 4** (Pinsker's Inequality)**.** $\delta(p_u^A, P_u^B) \leq \sqrt{D(P_u^A||p_u^B)}$

**Exercise 5.** *Use Pinsker's Inequality to prove that* $\mathbb{E}[Z_i] \leq \sqrt{v_i}$*. Furthermore, show that* $\mathbb{E}[\sum Z_i] \leq \sqrt{k\sum v_i} = \sqrt{kI}$

This concludes the proof of number of iterations less than $\sqrt{kI}$ and thus since each iteration is $\log k$ proves the BBCR theorem □

# 4 Amortized Communication / Direct Product Problems

Suppose we have the function $f : \mathcal{X}, \mathcal{Y} \to R$ and wish to compute $f^{\otimes n} : \mathcal{X}^n, \mathcal{Y}^n \to R^n$, where

$f^{\otimes n}(x_1, x_2, \ldots x_n, y_1, y_2, \ldots y_n) = (f(x_1, y_1), f(x_2, y_2), \ldots f(x_n, y_n))$. How does $CC(f^{\otimes n})$ compare with $CC(f)$? We can just take the protocol for $f$ and execute it $n$ times, to get $CC(f^{\otimes n}) \leq CC(f)n$.

Long time open problem: Is it true that $CC(f^{\otimes n}) \geq \Omega(CC(f)n)$, up to constant factors?

[BBCR] show that $CC_\varepsilon^n(f) \geq \Omega(CC_\varepsilon(f)\sqrt{n})$

**Theorem 6** (BR)**.** $IC_\varepsilon(f^{\otimes n}) = IC_\varepsilon(f)n$.

Suppose we have some protocol $\pi$ for $f^{\otimes n}$ with $CC(\pi) = C$. This implies $CC(\pi) = C$ and $IC(\pi) \leq C \Rightarrow f$ has protocol with $IC \leq \frac{C}{n}$, and compressing gives a protocol for $f$ with $CC = O(\frac{C}{\sqrt{n}})$.

This theorem is interesting: if we have a problem, solving its $n$-fold takes a scaled by n amount of communicating.

## 4.1 Let's talk about errors!

**Definition 7.** We say some protocol $\pi^n$ solves $f^{\otimes n}$ with error $\varepsilon$ if $\forall i$ we have $\Pr[(\pi^n)_i \neq (f^{\otimes n})_i] \leq \varepsilon$

**Definition 8.** $CC_\varepsilon^n(f) = \min_\pi\{CC(\pi)\}$ s.t. $\pi$ solves $f^{\otimes n}$ with error $\varepsilon$

**Definition 9.** $IC_\varepsilon^n(f) = \min_\pi\{IC(\pi)\}$ s.t. $\pi$ solves $f^{\otimes n}$ with error $\varepsilon$

$CC_\varepsilon^n(f)$ is bounded strictly above by $CC_\varepsilon(f^{\otimes n})$.

# 5 Next Time

[BR]: $CC_\varepsilon^n(f) = (1 + o(1))IC_\varepsilon^n(f)$

For the interactive correlated sampling problem, $CC(ICS) = D(P||Q) + O(\sqrt{D(P||Q)} + \log\frac{1}{\varepsilon})$ for input distribution $P$ to Alice and $Q$ to Bob.

[GKR]: $\exists f_k$ s.t. $IC(f_k) \leq k$ and $CC(f_k) \geq 2^k$.

This establishes that the answer to $CC(f^{\otimes n}) \geq \Omega(CC(f)n)$ is NO.