CS 229r Information Theory in Computer Science March 10, 2016

Lecture 14

Lecturer: Guannan Qu *Scribe: Sahana Vasudevan*

1 Goal of Today's Class

We will show that Information = Amortized Comumication Complexity.

That is, let $f(x,y)$ be a function, μ be a distribution on the the inputs to f, and ρ be the error probability. We define the Amortized Communication Complexity of f to be $\lim_{n\to\infty} \frac{1}{n} CC(f^{\otimes n})$. The goal is to show the following.

Theorem 1 $\lim_{n\to\infty} \frac{1}{n} CC(f^{\otimes n}) = IC^i(f)$.

We will denote $IC^i(f) = IC(f)$ since there is no external information cost here.

We showed in the previous lecture that $IC(f) \le \frac{1}{n} CC(f^{\otimes n}) \le CC(f)$. This result tells us that the first inequality becomes an equality, as $n \to \infty$.

The direct sum problem is the problem of determining the relationship between $\lim_{n\to\infty} \frac{1}{n} CC(f^{\otimes n})$ and $CC(f)$. Our result here reduces this problem to the one of determining the relationship between $IC(f)$ and $CC(f)$.

2 Sampling Protocol

We prove this theorem by introducing a protocol that satisfies the following properties. Suppose player A is given a distribution P and player B is given a distribution Q. We claim there exists a protocol such that at the end,

1. A outputs x drawn from P.
2. B outputs y s.t. for all x, $P(y = x|x) > 1 - \epsilon$.
3. Expected Communication $= D(P \parallel Q) + 5\sqrt{D(P \parallel Q)} + O(\log \frac{1}{\epsilon} + 1)$.

The protocol works as follows. Suppose the distributions P and Q are given over a universe U. The shared random tape can be interpreted as a sequence of $(x_i, p_i) \in U \times [0,1]$. We use this random tape to also get hash functions $h_i : U \to [0,1]$ so that $P(h_i(x) = h_i(y)) = 1/2$ for all $x \neq y$, i. The following are the steps of the protocol.

1. A picks a point (x, p) which is the first point in the public randomness lying below P; let i be the index of this point.
2. A sends $k = \lceil \frac{i}{|U|} \rceil$ to B using $1 + \lceil \log\log \frac{1}{\epsilon} \rceil$ bits. If k is too large i.e. A needs to use more bits to send k, then abort.
3. Start with $t = 0$. At the tth iteration, the following occurs.
   (a) A sends $h_j(x_i)$ up to $j \le S(t) = (t+1)^2 + \log \frac{1}{\epsilon} + 1$.
   (b) B finds if there exists r such that $(x_r, p_r) \in C_t Q$ (where $C_t = 2^{t^2}$) and also $h_j(x_r) = h)j(x_i) \forall j \le S(t)$. If r exists, output x_r. Otherwise, t increments by 1 and B outputs failure.

We claim this protocol satisfies the desired properties. First, we bound the total communication. Note that step 3 terminates when $t^2 \geq \log P(x_i)/Q(x_i)$. Let $T = \lceil \sqrt{\log P(x_i)/Q(x_i)} \rceil$. Then by the Tth iteration, A will have sent at most $S(T)$ bits and B at most $T+1$ bits, so the total communication in step 3 is at most $S(T)+T+1$. Taking expectation gives the desired bound on the expected communication.

Next we analyze the error probability. We note that disagreement occurs if there exists $r \neq i$ such that $(x_r, p_r) \in C_t Q$ and $h_j(x_r) = h_j(x_i)$ for all $j \leq S(t)$. This happens with probability $\frac{C_t}{|U|} 2^{-S(t)}$ for a particular r. Taking the union bound among all r, we see the total error probability is less than $\epsilon 2^{-t}$.

3 Correlated Pointer Jumping

We define a related problem called Correlated Pointer Jumping as follows. The input is a rooted tree where each non-leaf node is owned by A or B, each non-leaf node owned by a particular player has a set of children owned by the other, corresponding to each node v there are two distributions, T_v^A of children known to A and T_v^B of children known to B. We have a distribution on the tree obtained by sampling each child of the node according to the distribution given by the owner of the parent node. The goal is to sample the leaves of the tree according to this distribution.

Now given a public protocol π with inputs X, Y and randomness R, fixing the inputs x, y and randomness r, we get an instance of correlated pointer jumping as follows. The tree is the corresponding protocol tree, and the distributions T_v^A and T_v^B are defined as follows. Suppose v is a node of depth i and is owned by A. Let T_v be the random variable corresponding to which child of v is chosen; then $T_v^A = T_v | X = x, \pi_{<i}(X,Y) = rv$ and $T_v^A = T_v | Y = y, \pi_{<i}(X,Y) = rv$.

Let k be the depth of the protocol tree. We can use the method in the previous section to solve this problem by repeatedly running the previous protocol.

If the path sampled is $T = (v_0, v_1, ..., v_k)$, then we have that the communication is

$$\sum_{i=0}^{k} D(P_{T_{v_i}^A} \parallel P_{T_{v_i}^B}) + 5\sqrt{D(P_{T_{v_i}^A} \parallel P_{T_{v_i}^B})} + O(\log \frac{1}{\epsilon} + 1)$$

$$= \sum_{i=0}^{k} (I(T_{v_i}^A | T_{v_i}^B) + I(T_{v_i}^B | T_{v_i}^A)) + 5\sqrt{I(T_{v_i}^A | T_{v_i}^B) + I(T_{v_i}^B | T_{v_i}^A)} + O(\log \frac{1}{\epsilon} + 1)$$

$$\leq IC(T) + 5\sqrt{kIC(T)} + kO(\log \frac{1}{\epsilon} + 1)$$

where the last inequality is by Cauchy-Schwartz.

4 Proof of Theorem 1

It suffices to show that for any $\delta > 0$, for sufficiently large n we have $CC(f^{\otimes n}) < IC(f) + \delta/2$. The idea is to take a protocol π that computes f with error $< \alpha$ with respect to μ, for some $\alpha < \rho$. Let π^n be the protocol that takes inputs in X^n, Y^n and parallely runs them. We simulate π^n as described in the previous section with error $(\rho - \alpha)/2$, and truncate after $IC(\pi^n) + 5\sqrt{CC(\pi) IC(\pi^n)} + CC(\pi) O(\log \frac{1}{\epsilon} + 1) = nIC(\pi) + 5\sqrt{nCC(\pi) IC(\pi)} + CC(\pi) O(\log \frac{1}{\epsilon} + 1)$ bits. For sufficiently large n, this communication is at most $n(IC(\pi) + (\delta/2))$, and it can be shown that the error is at most ρ. This shows $CC(f^{\otimes n}) < IC(f) + \delta/2$ for sufficiently large n, as desired. This completes the proof of Theorem 1.

5 References

Mark Braverman and Anup Rao. Information equals amortized communication. CoRR, abs/1106.3595, 2011.