COMP760, SUMMARY OF LECTURE 21.

HAMED HATAMI

1. Set Disjointness

In the previous lecture we proved the following theorem.

Theorem 1. *We have*

$$\lim_{\gamma\to 0}\max_{\mu:\mu(11)<\gamma}\mathrm{IC}_\mu(\mathrm{AND},\mu,0)=\max_{\mu:\mu(11)=0}\mathrm{IC}_\mu(\mathrm{AND})=0.482702\ldots.$$

Recall that we showed $\mathrm{IC}_\mu(f,\mu,\epsilon)$ is continuous with respect to ϵ at every point $\epsilon\in(0,1]$. In fact $\mathrm{IC}_\mu(f,\mu,\epsilon)$ is continuous at $\epsilon=0$ too. However the proof is more involved, and we refer the reader to [BGPW13] for a proof.

Theorem 2 ([BGPW13])**.** *For every function f and measure μ, and every $\delta\in[0,1]$ we have*

$$\lim_{\epsilon\to\delta}\mathrm{IC}_\mu(f,\mu,\epsilon)=\mathrm{IC}_\mu(f,\mu,\delta).$$

One can deduce the following corollary from Theorem 1 and continuity.

Corollary 3. *We have*

$$\lim_{\epsilon\to 0}\max_{\mu:\mu(11)=0}\mathrm{IC}_\mu(\mathrm{AND},\epsilon)=0.482702\ldots.$$

Proof. The "$\leq$" direction is obvious from Theorem 1. It remains to prove the "$\geq$" direction. The proof of Theorem 2 shows that for every $\epsilon>0$, we have

$$\mathrm{IC}_\mu(\mathrm{AND},\mu,0)-\delta_\epsilon\leq\mathrm{IC}_\mu(\mathrm{AND},\mu,\epsilon)\leq\mathrm{IC}_\mu(\mathrm{AND},\epsilon),$$

with $\lim_{\epsilon\to0}\delta_\epsilon=0$. Note that δ_ϵ does not depend on μ. Now let us take the maximum over all μ with $\mu(11)\leq\gamma$ where $\gamma>0$. We have

$$\max_{\mu:\mu(11)\leq\gamma}\mathrm{IC}_\mu(\mathrm{AND},\mu,0)-\delta_\epsilon\leq\max_{\mu:\mu(11)\leq\gamma}\mathrm{IC}_\mu(\mathrm{AND},\epsilon).$$

Theorem 1 shows that taking the limit of $\gamma\to0$, the left hand side converges to $(0.482\ldots)-\delta_\epsilon$. Hence we obtain

$$(0.482\ldots)-\delta_\epsilon\leq\max_{\mu:\mu(11)\leq\gamma}\mathrm{IC}_\mu(\mathrm{AND},\epsilon).$$

Taking the limit $\epsilon\to0$ completes the proof. □

In this lecture we will show how one can use this corollary to determine the exact asymptotics of the randomized communication complexity of the DISJ problem. The AND function comes from the fact that

$$\overline{\mathrm{DISJ}}(X,Y)=\vee_{i=1}^n(x_i\wedge y_i).$$

Theorem 4. *We have*

$$\lim_{\epsilon\to0}\frac{R_\epsilon(\mathrm{DISJ}_n)}{n}=0.482702\ldots.$$

1.1. **Proof of Theorem 4: the lower-bound.** Consider a protocol π_ϵ that computes $\overline{\text{DISJ}}$ with error probability at most $\epsilon > 0$ and has the optimal communication cost $\text{CC}(\pi_\epsilon) = R_\epsilon(\text{DISJ})$. Consider any measure $\nu \in \Delta(\{0,1\} \times \{0,1\})$ with $\nu(11) = 0$. We will use π_ϵ to obtain a protocol τ_ϵ that solves AND with error probability at most ϵ, and with information cost at most $\text{CC}(\pi_\epsilon)/n$ under ν. The idea is to use the same approach that we used to prove

$$\text{IC}_{\mu^n}(T^n) = n\text{IC}_\mu(T).$$

Namely, this protocol τ_ϵ is obtained by restricting π_ϵ to a single random coordinate where the other coordinates are sampled randomly (some parts privately and some publicly) according to ν. See Figure 1. As we have already shown earlier in the course:

$$\text{IC}_\nu(\tau_\epsilon) = \frac{\text{IC}_{\nu^n}(\pi_\epsilon)}{n}.$$

Note further that since $\nu(11) = 0$, if (X_i, Y_i) are sampled according to ν for $i \in [n] \setminus \{j\}$ (as they are in τ_ϵ), then with probability 1 we have

$$\pi_\epsilon((X_1, \ldots, X_{j-1}, x, X_{j+1}, \ldots, X_n), (Y_1, \ldots, Y_{j-1}, y, Y_{j+1}, \ldots, Y_n)) = x \wedge y.$$

It follows that for every $(x, y) \in \{0,1\} \times \{0,1\}$ we have

$$\Pr[\tau_\epsilon(x, y) \neq (x \wedge y)] \leq \epsilon.$$

Hence

$$\frac{R_\epsilon(\text{DISJ}_n)}{n} = \frac{\text{CC}(\pi_\epsilon)}{n} \geq \text{IC}_\nu(\text{AND}, \epsilon).$$

Taking the limit as $\epsilon \to 0$, we obtain

$$\lim_{\epsilon \to 0} \frac{R_\epsilon(\text{DISJ}_n)}{n} \geq \lim_{\epsilon \to 0} \max_{\nu : \nu(11)=0} \text{IC}_\nu(\text{AND}, \epsilon) = 0.482702\ldots.$$

FIGURE 1. The protocol τ_ϵ that solves AND using π_ϵ that solves $\overline{\text{DISJ}}_n$.

- On input (x, y) for the AND function, Alice and Bob publicly choose:
  - $j \in \{1, \ldots, n\}$ uniformly at random, and set $X_j = x$ and $Y_j = y$ individually.
  - $X_1, \ldots, X_{j-1}$ independently, each according to ν_x.
  - $Y_{j+1}, \ldots, Y_n$ independently, each according to ν_y.
- Alice privately chooses $X_{j+1}, \ldots, X_n$ according to ν conditioned on the values of $Y_{j+1}, \ldots, Y_n$.
- Bob privately chooses $Y_1, \ldots, Y_{j-1}$ according to ν conditioned on the values of $X_1, \ldots, X_{j-1}$.
- They run the protocol π_ϵ on (X, Y) with $X = (X_1, \ldots, X_n)$ and $Y = (Y_1, \ldots, Y_n)$, and output the j-th coordinate.

1.2. **Proof of Theorem 4: the upper-bound.** Consider a measure (not necessarily a product measure) μ on $\{0,1\}^n \times \{0,1\}^n$. First note that if $|\{i : X_i = Y_i\}|$ is typically large when (X, Y) are sampled according to μ, then the average case communication complexity of DISJ is small under this distribution. Indeed if Alice and Bob exchange some random coordinates, then soon they will discover an intersection. Hence the difficult measures are the ones whose coordinate marginals μ_i satisfy $\mu_i(11) = o(1)$ for almost all i. First we will construct a protocol that has zero error and low information cost (but has large communication cost).

Theorem 5. *There is a protocol τ that solves* DISJ_n *correctly on all inputs and for every distribution μ satisfies*

$$\mathrm{IC}_\mu(\pi) \leq (0.482702\ldots)n + o(n).$$

Proof. Consider the protocol in Figure 2.

FIGURE 2. A protocol τ with small information cost for DISJ.

- For $i = 1, \ldots, M := o(n)$ do
- Publicly pick a random coordinate j and exchange the inputs bits on this coordinate. If they are both 1, output 1 and terminate.
- For every coordinate run $\pi_\wedge$ on that coordinate and if it outputs 1, output 1 and terminate.
- If not terminated yet, output 0.

Let E be the random variable that is 1 if they find an intersection in the first M coordinates, and 0 otherwise. Let $\Pi_1\Pi_2$ be the transcript where Π_1 corresponds to the first phase where the M coordinates are probed, and Π_2 the rest of the protocol. Since Π_1 determines E, and conditioned on EXY, Pi_2 is independent of Π_1, we have

$$\begin{aligned} I(\Pi; X|Y) &= I(\Pi_1\Pi_2; X|Y) = I(\Pi_1; X|Y) + I(\Pi_2; X|\Pi_1 Y) = I(\Pi_1; X|Y) + I(\Pi_2; X|\Pi_1 EY) \\ &\leq I(\Pi_1; X|Y) + I(\Pi_2; X|EY) \leq 2M + I(\Pi_2; X|EY) \\ &= 2M + \Pr[E = 0] I(\Pi_2; X|Y, E = 0) \leq 2M + \Pr[E = 0](2n). \end{aligned}$$

This shows that the theorem holds if $\Pr[E = 0] = o(1)$. Hence we can assume $\Pr[E = 0] = \Omega(1)$. By taking M to be sufficiently large we can guarantee that for sufficiently large n we have

$$\Pr[(|X \cap Y| \geq \sqrt{n}) \wedge (E = 0)] \leq \frac{1}{n^2},$$

and hence

$$\Pr[(|X \cap Y| \geq \sqrt{n}) \mid (E = 0)] \leq \frac{\Omega(1)}{n} \leq \frac{1}{n},$$

for sufficiently large n. Consequently

$$\mathbb{E}[|X \cap Y| \mid E = 0] \leq \sqrt{n}.$$

The following theorem shows that the information cost of a protocol τ that runs a protocol π independently on many copies chosen according to the joint distribution ν is less or equal than the sum of the information costs of the protocol π on the marginals $\nu_1, \ldots, \nu_n$.

Theorem 6 (See [Bra12, Theorem 4.2]). *Let ν be a distribution on $\{0,1\}^n \times \{0,1\}^n$ and let $\nu_1, \ldots, \nu_n$ be marginals of ν on the coordinates $1, \ldots, n$. Let π be a protocol on $\{0,1\} \times \{0,1\}$ and let π^n be the protocol that runs π on all the coordinates. Then*

$$\mathrm{IC}_{\nu^n}(\pi^n) \leq \mathrm{IC}_{\nu_1}(\pi) + \ldots + \mathrm{IC}_{\nu_n}(\pi).$$

Let $\mu_1, \ldots, \mu_n$ be marginals of $\mu|_{E=0}$ on the coordinates $1, \ldots, n$. Let $\epsilon_i := \Pr[X_i = Y_i = 1 | E = 0] = \mu_i(11)$ for $i = 1, \ldots, n$. Note

$$\epsilon_1 + \ldots + \epsilon_n = \mathbb{E}[|X \cap Y| \mid E = 0] \leq \sqrt{n}.$$

Let J be the set of coordinates with $\epsilon_i \geq \frac{1}{n^{1/4}}$. Then $|J| \leq \frac{\sqrt{n}}{n^{1/4}} = n^{3/4}$. Thus we conclude that

$$\begin{aligned} \mathrm{IC}_\mu(\tau) &\leq 4M + \mathrm{IC}_{\mu_1}(\pi_\wedge) + \ldots + \mathrm{IC}_{\mu_n}(\pi_\wedge) \leq 4M + 2n^{3/4} + \sum_{i \notin J} \mathrm{IC}_{\mu_i}(\pi_\wedge) \\ &= (0.482702\ldots)n + o(n), \end{aligned}$$

by Theorem 1 as $\mu_i(11) = o(1)$ for $i \notin J$. □

Finally we prove the upper-bound (in Theorem 4) on the randomized communication complexity of DISJ. The proof will use a prior free version of "information equals amortized communication". To define the prior-free information cost we need the following Min-Max theorem. Roughly speaking it says that if you are given a convex and compact set of measures, then there is a sequence of protocols that converges to the optimal information cost (on the hardest measure) for all of these measures *simultaneously*.

Theorem 7 (An information theoretic Min-Max theorem [Bra12]). *Let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ be a function, and let $K \subseteq \Delta(\mathcal{X} \times \mathcal{Y})$ be closed and convex. Then*

$$\inf_\pi \max_{\mu \in K} \mathrm{IC}_\mu(\pi) = \max_{\mu \in K} \inf_\pi \mathrm{IC}_\mu(\pi),$$

where both infimums are over protocols π that compute f correctly on all inputs.

Then for a function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ we can define

$$\mathrm{IC}(f) := \inf_\pi \max_{\mu \in \Delta} \mathrm{IC}_\mu(\pi) = \max_{\mu \in \Delta} \inf_\pi \mathrm{IC}_\mu(\pi).$$

Braverman [Bra12] proved the following "information equals amortized communication" theorem.

Theorem 8 ([Bra12]). *Let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ be a function, and let $\mathrm{IC}(f) = I$. Then for every $\delta_1, \delta_2 > 0$, and sufficiently large $N > N_{\delta_1,\delta_2}(f)$, there exists a protocol π_N that computes N copies of f, and has communication cost $\leq N \cdot I \cdot (1 + \delta_1)$, and answers correctly on all coordinates except with probability δ_2.*

Finally we will state the proof of the upper-bound in Theorem 4.

The upper-bound in Theorem 4. Theorem 5 shows that we can solve disjointness with low information cost. How can we convert this to low communication cost? The trick is to use the so called "self-reducibility" properties of the disjointness function to treat solving one instance of it as solving many smaller instances of it in parallel. Then one can apply "information equals amortized communication" to finish the proof.

Consider a sufficiently large m and N and let $n = N \cdot m$, and $\delta = 1/m$. In particular we assume $N \geq N_{\delta,\delta}(\mathrm{DISJ})$. Theorem 5 shows that $\mathrm{IC}(\mathrm{DISJ}_m) \leq (0.4827\ldots)m + o(m)$. By Theorem 8, there

is a protocol π_N that solves N instances of DISJ_m with communication at most

$$\begin{aligned} N \cdot \mathrm{IC}(\mathrm{DISJ}_n) \cdot (1+\delta) &= N((0.4827\ldots)m + o(m))\,(1+\delta) = (0.4827\ldots)n + o(N) \\ &= (0.4827\ldots)n + o(n), \end{aligned}$$

and with error at most δ. Then the protocol in Figure 3 solves DISJ with error at most δ and has communication cost at most $(0.4827\ldots)n + o(n)$.

FIGURE 3. A protocol with small communication cost for DISJ.

- Divide $n = N \cdot m$ inputs into N blocks of size m.
- Run π_N on each block.
- Output 0 if π_N outputs 0 on some block.

□

REFERENCES

[BGPW13] Mark Braverman, Ankit Garg, Denis Pankratov, and Omri Weinstein, *From information to exact communication (extended abstract)*, STOC'13—Proceedings of the 2013 ACM Symposium on Theory of Computing, ACM, New York, 2013, pp. 151–160. MR 3210776

[Bra12] Mark Braverman, *Interactive information complexity*, STOC'12—Proceedings of the 2012 ACM Symposium on Theory of Computing, ACM, New York, 2012, pp. 505–524. MR 2961528

SCHOOL OF COMPUTER SCIENCE, MCGILL UNIVERSITY, MONTRÉAL, CANADA
E-mail address: hatami@cs.mcgill.ca