COMP760, SUMMARY OF LECTURE 14.

HAMED HATAMI

1. Introduction to information complexity

So far, in this course, we have been focused on the communication cost of protocols, and given various communication tasks, we studied the minimum communication cost required to perform those tasks. The information theoretic tools introduced in the past two lectures enable us to introduce another natural measure of complexity. Roughly speaking, given a protocol and a distribution on the inputs of Alice and Bob, the information cost of this protocol is the amount of information that the communication between Alice and Bob reveals about their inputs. Note that a protocol with a very high communication cost might still have a small information cost if the communication between Alice and Bob does not reveal much about their inputs. Naturally, we will be interested in finding the smallest information cost required to perform a given communication task.

The concept of information complexity is deeply connected to communication complexity. Recall how entropy captured the amount of information in a random variable, and how this quantity gave us the exact asymptotic of the transmission cost of many independent copies of X. That is the simplest setting of communication where there is a one-way channel and Alice wants to transmit her data to Bob. The general setting of communication complexity is more complicated as Alice and Bob are allowed to interact. However as the recent results in this area have demonstrated, similar to the way that the information content of a random variable gives the asymptotics of the transmission cost, information complexity provides valuable information about the communication complexity, and specially in the asymptotic case where Alice and Bob are performing many instances of a given communication task.

1.1. **The basic setting.** The setting is the same as the two player setting in communication complexity, where Alice and Bob (having infinite computational power) want to mutually compute a function $f : X \times Y \to \{0,1\}$. To be able to measure information, we also need to assume that there is a prior distribution μ on $X \times Y$. Earlier in the course, we discussed two different models of randomized communication protocols, the public coin and private coin protocols. For the purpose of communication complexity, once we allow public randomness, it makes no difference whether we permit the players to have private random strings or not. This is because the private random strings can be simulated by parts of the public random string, which is infinite. However, for information complexity, it is crucial to consider protocols that permit *both* private and public randomness. For example, consider a protocol in which Alice sends $x \oplus r$ to Bob where r is chosen uniformly at random from $\{0,1\}^n$. Note that if r is private to Alice, then the communication of $x \oplus r$ does not reveal any information about x, while if r is known publicly, then $x \oplus r$ completely reveals x. Hence we will consider the setting where both public and private random strings are present. We shall also consider the deterministic case, where there is not randomness in the protocol, or the restricted models where only one of public or private randomness is permitted.

- A function $f : X \times Y \to \{0,1\}$, and a probability distribution μ on $X \times Y$.
- A public random string R (visible to everybody), and private random strings R_A (known to Alice), and R_B (known to Bob).

- Alice receives $x \in X$ and Bob receives $y \in Y$, where (x, y) is sampled randomly according to the distribution μ.

1.2. **Allowing long messages in one round.** One of the greatest achievements of information complexity has been its success [BGPW13] in determining the exact asymptotics of randomized communication *complexity* of functions such as disjointness

$$\lim_{\epsilon \to \infty} \frac{R_\epsilon(\mathrm{DISJ}_n)}{n} = 0.4827.$$

Prior to the discovery of these techniques, proving the lower-bound $R_\epsilon(\mathrm{DISJ}_n) = \Omega(n)$ was already a challenging problem and it took several works to find a short proof for this fact [Raz92]. To be able to talk about the communication complexity with this precision we shall sometimes need to use a more economical definition of a protocol. Since we had only been interested in the big-O asymptotics of communication complexity, we simplified the model and assumed that Alice and Bob alternatively speak and each time they send a bit to the other party. However this might unnecessarily double the cost of communication, as there might be an optimal protocol in which one of the parties remains quiet in many of the rounds. So in our revised model we will let each one of the parties, Alice and Bob, to send a message longer than one bit if necessary. Also in order for the other party to know that the message is finished we will assume that at each round the messages that a party can possibly send are prefix-free. That is none of the messages is the prefix (i.e. the start) of another possible message in that round. With this assumption, once the other party receives the last bit of the message, he knows that the message is finished (as at this stage, there is no possible message that can be obtained from this by possibly adding more bits), and does not need to receive an "over" message to confirm that the message is over.

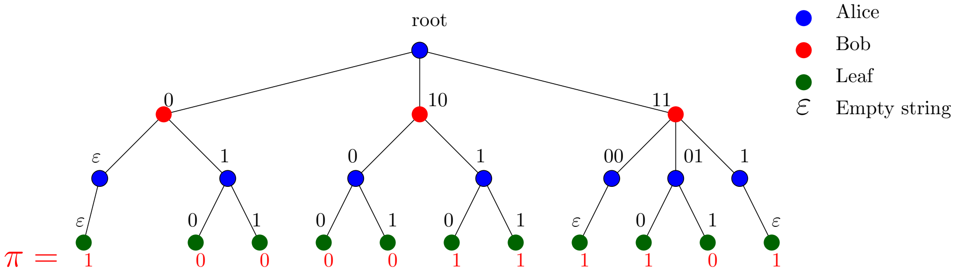
- A *private coin* protocol π is a tree where every internal node has an *owner* Alice or Bob.
- The owners alternate, that is the owner of every node is different from the owner of its parent.
- Every node of the tree, except the root, is labeled with a finite string in $\{0,1\}^{\mathbb{N}}$, such that the set M_v of the labels on the children of an internal node v is prefix-free.
- For every internal node v owned by Alice, there is a function $a_v : (x, R_A) \to M_v$. Here $a_v(x, R_A)$ is the message that will be sent by Alice at that stage of the execution of the protocol, and then the protocol will move the the corresponding child.
- For every internal node u owned by Bob, there is a function $b_u : (y, R_B) \to M_u$. Similarly $b_u(y, R_B)$ is the message that will be sent by Alice at that stage of the execution of the protocol, and then the protocol will move the the corresponding child.
- Every leaf is labeled with an "output" value. Note that they do not need to communicate this value as part of the communication. Once the protocol researches the leaf, they both have agreed on an output value.

We often assume that every leaf in the protocol is at the same depth. We can do this since if some leaf is at depth less than the maximum, we can modify the protocol by adding dummy nodes which are always picked with probability 1, until all leaves are at the same depth.

Note that the above model is a generalization of our original model where Alice and Bob were alternatively sending one bit at each round as then $M_v = \{0,1\}$ for every internal node, and it is prefix-free.

The reason that we do not require the parties to communicate the output value is that, the output might be very big, and outputting it can be very costly as we will be working with functions $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ where $\mathcal{Z}$ can be a large set.

FIGURE 1. An example of a protocol tree



Now we can define a protocol with both public coin and private coin as a distribution on private coins, run by first using shared randomness to sample an index r, and then running the corresponding private coin protocol π_r.

1.3. **The transcript.** The transcript of the protocol, denoted by $\Pi(x, y, R_A, R_B, R)$ is the concatenation of all the bits exchanged during the execution of the protocol together with the public random string R. It is basically all the information that is visible to both parties (and possibly to an external observer). We denote by Π_t the transcript of the t-th round of the communication. This is the message sent at that round. We also denote by $\Pi_0 = R$ the public randomness, so that

$$\Pi = \Pi_0\Pi_1 \dots \Pi_\ell,$$

where ℓ is the number of rounds. We shall also denote by $\Pi_{\leq t}$ the transcript of the communication from the beginning to the end of round t:

$$\Pi_{\leq t} = \Pi_0\Pi_1 \dots \Pi_t,$$

1.4. **Information cost.** We defined the *communication cost* of a protocol as the maximum number of bits that is exchanged by that protocol on all inputs and with all randomness:

$$\mathrm{CC}(\pi) = \max_{\substack{R_A, R_B, R\\ xy}} (\# \text{ of bits exchanged}).$$

We defined the *average communication cost* of a protocol as the maximum expected number of bits that is exchanged by that protocol on all inputs:

$$\mathrm{CC}^{\mathrm{avg}}(\pi) = \max_{xy} \mathbb{E}_{R_A, R_B, R}[\# \text{ of bits exchanged}].$$

We defined the *average communication cost* of a protocol with respect to the measure μ as the expected number of bits that is exchanged by that protocol when the inputs are drawn from the distribution μ.

$$\mathrm{CC}_\mu(\pi) = \mathbb{E}_{\substack{XY \sim \mu\\ R_A, R_B, R}} [\# \text{ of bits exchanged}].$$

Obviously

$$\mathrm{CC}_\mu(\pi) \leq \mathrm{CC}^{\mathrm{avg}}(\pi) \leq \mathrm{CC}(\pi).$$

Finally we are ready to formally define the information cost of a protocol. Maybe the most natural attempt would be to define the information cost as the amount of information that is revealed about the inputs X and Y to an external observer who sees that communication and the public randomness. This is known as the *external information cost* and is formally defined as the mutual information between XY and the transcript [1] of the protocol

$$\mathrm{IC}^{\mathrm{ext}}_{\mu}(\pi) = I(XY;\Pi).$$

While this notion is interesting and useful, it turns out there is another way of defining information cost that has nicer properties. This is called the *internal information cost* or just the *information cost* for short, and it corresponds to the amount of information that Alice and Bob learn about each other's inputs from the communication. Note that Bob knows Y and R_B, and thus what he learns about X *from the communication* can be measured by $I(X;\Pi|YR_B)$ and similarly what Alice learns about Y from the communication can measured by $I(Y;\Pi|XR_A)$. As we shall see below in Proposition 1, conditioning on private randomness does not affect these quantities: $I(X;\Pi|Y,R_B) = I(X;\Pi|Y)$ and $I(Y;\Pi|X,R_A) = I(Y;\Pi|X)$. Hence we define the information cost as

$$\mathrm{IC}_{\mu}(\pi) = I(X;\Pi|Y) + I(Y;\Pi|X).$$

The above definition is fairly recent and it is due to Barak, Braverman, Chen and Rao [BBCR10] from 2009.

1.5. **Transcripts and conditioning on them .** Before proving the basic results about information cost, let us list some useful facts that will be used frequently. Consider a specific transcript Π of a protocol π. Note that the transcript corresponds to combinatorial rectangle $\mathcal{X}_\Pi \times \mathcal{Y}_\Pi$, where $\mathcal{X}_\Pi$ is a set of possible values for XR_A, and $\mathcal{Y}_\Pi$ is a set of possible values for YR_B. Hence conditioning on Π means conditioning on the event

$$(R = \Pi_0, XR_A \in \mathcal{X}_\Pi, YR_B \in \mathcal{Y}_\Pi).$$

This is a very useful fact. For example it immediately implies the following conditional independences:

$$I(XR_A;R_B|Y\Pi) = I(R_A;YR_B|X\Pi) = 0,$$

and

$$I(R_A;R_B|XY\Pi) = I(R_A;R_B|X\Pi) = I(R_A;R_B|Y\Pi) = 0.$$

1.6. **Conditioning on Private randomness does not matter.**

Proposition 1. *We have* $I(X;\Pi|YR_B) = I(X;\Pi|Y)$ *and* $I(Y;\Pi|XR_A) = I(Y;\Pi|X)$.

The proof will use the identity

$$I(A;B|Z) = I(A;B|ZC) - I(A;C|ZB) + I(A;C|Z), \tag{1}$$

from the last lecture.

Proof of proposition 1. We only prove $I(X;\Pi|YR_B) = I(X;\Pi|Y)$, as the proof of the other equality is identical. The proof is by induction. Suppose that the interaction has 2ℓ rounds and Alice speaks at odd rounds $1,3,5,\dots,2\ell-1$, and Bob speaks at even rounds $2,4,\dots,2\ell$. The statement that we want to prove is obvious for $\ell=0$, and our induction hypothesis says that

$$I(\Pi_{\le 2\ell-2};X|Y) = I(\Pi_{\le 2\ell-2};X|YR_B).$$

By chain rule

$$I(\Pi_{\le 2\ell};X|Y) = I(\Pi_{\le 2\ell-2};X|Y) + I(\Pi_{2\ell-1};X|Y\Pi_{\le 2\ell-2}) + I(\Pi_{2\ell};X|Y\Pi_{\le 2\ell-1}),$$

[1] Recall that Π contains the public random string R

and

$$I(\Pi_{\leq 2\ell}; X|YR_B) = I(\Pi_{\leq 2\ell-2}; X|YR_B) + I(\Pi_{2\ell-1}; X|YR_B\Pi_{\leq 2\ell-2}) + I(\Pi_{2\ell}; X|YR_B\Pi_{\leq 2\ell-1}).$$

The 3rd term is 0 in both identities by the discussion in Section 1.5 as Bob speaks at round 2ℓ and what he says is based on $YR_B\Pi_{\leq 2\ell-1}$. Hence, using the induction hypothesis, to establish the proposition it suffices to prove

$$I(\Pi_{2\ell-1}; X|Y\Pi_{\leq 2\ell-2}) = I(\Pi_{2\ell-1}; X|YR_B\Pi_{\leq 2\ell-2}). \tag{2}$$

By the identity (1) we have

$$I(\Pi_{2\ell-1}; X|Y\Pi_{\leq 2\ell-2}) = I(\Pi_{2\ell-1}; X|Y\Pi_{\leq 2\ell-2}R_B) - I(\Pi_{2\ell-1}; R_B|Y\Pi_{\leq 2\ell-2}X) + I(\Pi_{2\ell-1}; R_B|Y\Pi_{\leq 2\ell-2}).$$

The last two terms are equal to 0 by the discussion in Section 1.5 as Alice speaks at round $2\ell - 1$, and her decision is based on $\Pi_{\leq 2\ell-2}XR_A$. □

1.7. **Information cost is the average over public randomness.** Consider a protocol π, and denote by π_r the protocol obtained by fixing the public randomness to a string r.

Theorem 2. *We have*

$$\mathrm{IC}_\mu(\pi) = \mathbb{E}_r \mathrm{IC}_\mu(\pi_r).$$

Proof. We have

$$\begin{aligned} \mathrm{IC}_\mu(\pi) &= I(\Pi; X|Y) + I(\Pi; Y|X) \\ &= I(R; X|Y) + I(\Pi; X|YR) + I(R; Y|X) + I(\Pi; Y|XR) \\ &= I(\Pi; X|YR) + I(\Pi; Y|XR) = \mathbb{E}_r I(\Pi; X \mid Y, [R=r]) + I(\Pi; Y \mid X, [R=r]) \\ &= \mathbb{E}_r \mathrm{IC}(\pi_r, \mu), \end{aligned}$$

where $[R = r]$ denotes the event that $R = r$.

□

1.8. **External information cost is larger than the information cost.** The following theorem shows that $\mathrm{IC}^{\mathrm{ext}} \geq \mathrm{IC}$. The intuitive reason behind this fact is that the external observer does not know either of X or Y, so she can learn more new information about the inputs than Alice and Bob who already have some information about the other person's input from the possible correlation between X and Y. Indeed if X and Y are independent and thus there is no such correlation between them (which is equivalent to μ being a product distribution), then $\mathrm{IC}^{\mathrm{ext}}$ and IC are equal.

Theorem 3. *We have*

$$\mathrm{IC}_\mu(\pi) \leq \mathrm{IC}^{\mathrm{ext}}_\mu(\pi).$$

Proof. Suppose without loss of generality that the interaction has 2ℓ rounds and Alice speaks at odd rounds, and Bob speaks at even rounds. By induction hypothesis the statement is true for $2\ell - 1$ rounds, and combining this with the chain rule $\mathrm{IC}^{\mathrm{ext}}_\mu(\pi_{2\ell-1}) = I(XY; \Pi_{2\ell-1}) = I(Y; \Pi_{2\ell-1}) + I(X; \Pi_{2\ell-1}|Y)$ shows that

$$I(Y; \Pi_{\leq 2\ell-1}) \geq I(Y; \Pi_{\leq 2\ell-1}|X). \tag{3}$$

Similarly to establish the induction hypothesis for the 2ℓ rounds, it suffices to prove

$$I(Y; \Pi) \geq I(Y; \Pi|X).$$

By (3), we have

$$\begin{aligned} I(Y;\Pi) &= I(Y;\Pi_{\le 2\ell-1}) + I(Y;\Pi_{2\ell}|\Pi_{\le 2\ell-1}) \\ &\ge I(Y;\Pi_{\le 2\ell-1}|X) + I(Y;\Pi_{2\ell}|\Pi_{\le 2\ell-1}) \\ &= I(Y;\Pi_{\le 2\ell-1}|X) + I(Y;\Pi_{2\ell}|\Pi_{\le 2\ell-1}X) - I(X;\Pi_{2\ell}|\Pi_{\le 2\ell-1}Y) + I(X;\Pi_{2\ell}|\Pi_{\le 2\ell-1}) \qquad \text{By (1)} \\ &\ge I(Y;\Pi_{\le 2\ell-1}|X) + I(Y;\Pi_{2\ell}|\Pi_{\le 2\ell-1}X) - I(X;\Pi_{2\ell}|\Pi_{\le 2\ell-1}Y) \\ &= I(Y;\Pi_{\le 2\ell-1}|X) + I(Y;\Pi_{2\ell}|\Pi_{\le 2\ell-1}X) = I(Y,\Pi|X). \end{aligned}$$

□

1.9. **Information cost is bounded by communication cost.** The following theorem establishes the intuitive fact that the amount of information that Alice and Bob reveal about their input by their communication is bounded by the number of communicated bits.

Theorem 4. *We have*

$$\mathrm{IC}_\mu(\pi) \le \mathrm{IC}^{\mathrm{ext}}_\mu(\pi) \le \mathrm{CC}_\mu(\pi),$$

where $\mathrm{CC}_\mu(\pi)$ *denotes the average communication cost of* π *with respect to* μ*.*

Proof. We proved the first inequality in Theorem 3. To bound the external information cost by the average communication cost, note that

$$I(\Pi;XY) = I(\Pi_0;XY) + I(\Pi_{>0};XY|\Pi_0) = I(\Pi_{>0};XY|\Pi_0) \le H(\Pi_{>0}) \le \mathbb{E}[|\Pi_{>0}|] = \mathrm{CC}_\mu(\pi).$$

□

References

[BBCR10] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao, *How to compress interactive communication [extended abstract]*, STOC'10—Proceedings of the 2010 ACM International Symposium on Theory of Computing, ACM, New York, 2010, pp. 67–76. MR 2743255

[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

[Raz92] A. A. Razborov, *On the distributional complexity of disjointness*, Theoret. Comput. Sci. **106** (1992), no. 2, 385–390. MR 1192778 (93i:68095)

School of Computer Science, McGill University, Montréal, Canada
E-mail address: hatami@cs.mcgill.ca