COMP760, SUMMARY OF LECTURE 15.

HAMED HATAMI

1. Communication Tasks

In the past we discussed various communication problems: computing a function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ with no error

$$\Pr[f(x,y) \neq \pi(x,y)] = 0 \qquad \forall (x,y) \in \mathcal{X} \times \mathcal{Y};$$

computing a function with at most ϵ error on every input

$$\Pr[f(x,y) \neq \pi(x,y)] \leq \epsilon \qquad \forall (x,y) \in \mathcal{X} \times \mathcal{Y};$$

computing a function with expected error on ϵ measure of all inputs

$$\Pr_{XY \sim \mu}[\pi(X,Y) \neq f(X,Y)] \leq \epsilon.$$

We can think of these as communication tasks. Note that all the above definitions require something about the distribution of $(X, Y, \pi(X,Y))$. In the following definition we attempt to formally define a communication task. This is going to be useful when we discuss the direct sum theorems. However, the formalism of Definition 1 is not very essential for our purpose and the natural intuition that what a communication task will suffice.

Definition 1. *A* communication task *is a tuple $(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \Psi, \mu)$ where Ψ is a set of probability distributions on $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, and μ is a probability measure on $\mathcal{X} \times \mathcal{Y}$. We say that a communication protocol π performs this task if the distribution of $(X, Y, \pi(X,Y))$ belongs to Ψ if (X,Y) is sampled according to μ.*

Consider $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$, and let Ψ contain only the probability distribution that is the uniform distributed on $\{(x, y, f(x,y)) : (x,y) \in \mathcal{X} \times \mathcal{Y}\}$. Let μ be the uniform measure on $\mathcal{X} \times \mathcal{Y}$. Then a protocol π performs this task if it computes f with zero error. We denote this communication task by $[f]$.

Similarly, by letting μ be the uniform measure on $\mathcal{X} \times \mathcal{Y}$, we can define Ψ so that performing the task (Ψ, μ) would correspond to satisfying

$$\Pr[\pi(x,y) \neq f(x,y)] \leq \epsilon \qquad \forall (x,y) \in \mathcal{X} \times \mathcal{Y}.$$

We will denote this task by $[f, \epsilon]$.

Consider $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$, and let μ be a probability distribution on $\mathcal{X} \times \mathcal{Y}$. Then we can define Ψ so that performing the task (Ψ, μ) correspond to satisfying

$$\Pr_{XY \sim \mu}[\pi(X,Y) \neq f(X,Y)] \leq \epsilon.$$

This task is denoted by $[f, \mu, \epsilon]$.

Consider two tasks $T_1 = (\mathcal{X}_1, \mathcal{Y}_1, \mathcal{Z}_1, \Psi_1, \mu_1)$ and $T_2 = (\mathcal{X}_2, \mathcal{Y}_2, \mathcal{Z}_2, \Psi_2, \mu_2)$. Then $T_1 \times T_2$ corresponds to performing these two tasks at the same time. In other words a protocol

$$\pi_{r, r_A, r_B} : (\mathcal{X}_1 \times \mathcal{X}_2) \times (\mathcal{Y}_1 \times \mathcal{Y}_2) \to \mathcal{Z}_1 \times \mathcal{Z}_2$$

performs $T_1 \times T_2$ if when $X = (X_1, X_2)$ and $Y = (Y_1, Y_2)$ are sampled such that $X_1 Y_1 \sim \mu_1$ and $X_2 Y_2 \sim \mu_2$ are independent, then the distribution of $(X_1, Y_1, (\pi(X,Y))_1)$ belongs to Ψ_1 and the distribution of $(X_2, Y_2, (\pi(X,Y))_2)$ belongs to Ψ_2, where $(\pi(X,Y))_1$ and $(\pi(X,Y))_2$ denote respectively the first and the second coordinate of the output. In other words π is performing T_1 on the first coordinate and T_2 on the second coordinate.

It is a very interesting problem to study the asymptotics of the communication complexity of the task $T^n = T \times \ldots \times T$ as n goes to infinity. The related results are often categorized as direct product and direct sum theorems, and as we shall see information complexity has been very successful in answering some of the related questions.

1.1. Communication complexity and Information complexity of a function.

The *communication complexity* of a task T with respect to a measure μ is defined as

$$\mathrm{CC}_\mu(T) = \min_\pi \mathrm{CC}_\mu(\pi),$$

where the minimum is over all protocols π that perform T.

The *information complexity* of a task T with respect to a measure μ is defined as

$$\mathrm{IC}_\mu(T) = \inf_\pi \mathrm{IC}_\mu(\pi),$$

where the infimum is over all protocols π that perform T. As we will see later it is essential that we use infimum rather than minimum. In other words, there are communication tasks for which there is no protocol that achieves $\mathrm{IC}_\mu(T)$ while there is a sequence of protocols whose information cost converges to $\mathrm{IC}_\mu(T)$.

Recalling the tasks $[f]$, $[f,\epsilon]$ and $[f,\nu,\epsilon]$ we use the shorten notation

$$\begin{aligned}\mathrm{IC}_\mu(f) &= \mathrm{IC}_\mu([f])\\ \mathrm{IC}_\mu(f,\epsilon) &= \mathrm{IC}_\mu([f,\epsilon])\\ \mathrm{IC}_\mu(f,\nu,\epsilon) &= \mathrm{IC}_\mu([f,\nu,\epsilon]).\end{aligned}$$

Although the notation $\mathrm{IC}_\mu(f,\nu,\epsilon)$ permits us to measure the error with respect to ν while measuring the information cost with respect to a different measure μ, we will never use this freedom, and in all the examples that we will see $\nu = \mu$.

2. More basics about Information complexity

We start by making some basic observations about information complexity.

2.1. Continuity of information complexity.

The information complexities $\mathrm{IC}_\mu(f,\epsilon)$ and $\mathrm{IC}_\mu(f,\epsilon)$ are both continuous with respect to ϵ. The following theorem proves the continuity for $\epsilon \in (0,1]$. The continuity at 0 is more complicated and we shall prove it later in the course.

Theorem 2. *For each $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$, $\epsilon > 0$ and μ on $\mathcal{X} \times \mathcal{Y}$, we have*

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

and

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

Proof. Suppose $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$, and and consider a protocol π with information cost I, and error $\epsilon > 0$ (either worst case, or average error with respect to μ). Let τ be the following protocol

- With probability $1-\delta$ run π.
- With probability δ Alice and Bob exchange their inputs and compute $f(x,y)$.

The theorem follows as the new protocol has error $(1-\delta)\epsilon$ and information cost at most $I + \delta K$ where $K = \lceil \log|\mathcal{X}|\rceil + \lceil \log|\mathcal{Y}|\rceil$.

□

2.2. **Public randomness is unnecessary.** Note that when we are interested only in the information complexity we can assume that there is no public randomness. Indeed for example Alice can use some part of her private randomness to generate the public random string R and then send it to Bob, so that they will have a common random string. While this can be very costly in terms of communication, it has zero information cost, as the string R does not reveal any information about Alice's input. This observation is useful in the analysis of the information cost of generic protocols as it allows us to assume that the protocol does not use public randomness. Nevertheless we will still use public randomness in the description of the various protocols that we shall see in the sequel, as it is more natural to have a public random string from the beginning than having Alice generating it and sending it to Bob.

2.3. **The protocol tree and the probability of reaching a leaf .** Consider a protocol π, and fix the input (x, y), and let us also fix the public randomness to a fixed string $R = r$ (or alternatively we can assume that there is no public randomness). Consider an internal node u in the protocol tree at level k, and let us assume that Alice is the owner of u. Note that once we fix R, every internal node corresponds to a partial transcript, and every leaf corresponds to a full transcript. For example u corresponds to a partial transcript $\Pi_{\leq k-1}$.

Alice will use her private randomness to choose one of her possible messages (equivalently one of the children of u) randomly at this stage of the protocol, and this induces a probability distribution on the children of u. More precisely, for every child w of u, we have [1]

$$p_u^x(w) = \Pr_{R_A}[a_u(x, R_A, r) = w|u] = \Pr_{R_A}[\Pi_k(x, R_A, r) = w|R = r, \Pi_{\leq k-1} = u].$$

Similarly every internal node v at level k owned by Bob induces a probability distribution on the children of v.

$$p_v^y(w) = \Pr_{R_B}[b_v(y, R_B, r) = w|v] = \Pr_{R_B}[\Pi_k(y, R_B, r) = w|R = r, \Pi_{\leq k-1} = v].$$

Consider a full transcript t compatible with xyr (equivalently a leaf of the tree). What is the probability that t will be the transcript of the protocol for a given pair of inputs x, y? Let $(v_1, v_2, \ldots, v_\ell (= t))$ denote the unique path from the root to the leaf t, and let Π_{xy} be the random transcript generated on (x, y) (with fixed the public randomness r). Then we have

$$\Pr_{R_A,R_B}[\Pi_{xy} = t] = \left(\prod_{\substack{i\in[\ell-1] \\ \text{Alice owns } v_i}} p_{v_i}^x(v_{i+1}) \right) \times \left(\prod_{\substack{i\in[\ell-1] \\ \text{Bob owns } v_i}} p_{v_i}^y(v_{i+1}) \right) = p_A^x(t) p_B^y(t),$$

for two functions $p_A^x(\cdot)$ and $p_B^y(\cdot)$ known to Alice and Bob respectively. This product structure is very useful, and for example it implies the following rectangle property

$$\Pr[\Pi_{x_1y_1} = t]\Pr[\Pi_{x_2y_2} = t] = \Pr[\Pi_{x_1y_2} = t]\Pr[\Pi_{x_2y_1} = t].$$

Note that it also shows that given a particular t, Alice knows the probability $p_A^x(t)$ and Bob knows the probability $p_B^y(t)$, and for example Alice can send her number to Bob and then Bob can compute the probability of reaching the leaf t. We will use similar ideas frequently in the design of protocols that simulate a given protocol and whose communication cost depend on the information cost of the original protocol (compression).

[1] To be more precise we must write $p_u^{xr}(w)$, $p_v^{yr}(w)$, $p_A^{xr}(t), p_B^{yr}(t)$. However since we have fixed r, to simplify the notation, we drop it from the superscripts.

3. The equality function

In this section we show that under every distribution μ there is a protocol that solves the equality function with 0 error, and has information cost $O(1)$. On the other hand, we will show that there is a distribution μ such that every protocol that solves the equality function with 0 error has *external* information cost at least n. This shows that for the 0 error regime there can be a huge gap between the information complexity and the external information complexity of a function. Consider the following protocol for the equality function $\mathrm{EQ} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$.

```
• Alice and Bob publicly choose linearly independent $r_1, \ldots, r_n \in \mathbb{F}_2^n$ uniformly at random.
• For $i = 1$ to $n$ do:
•       Alice and Bob send $\langle X, r_i \rangle_{\mathbb{F}_2}$, and $\langle Y, r_i \rangle_{\mathbb{F}_2}$.
•       If they are *not* equal they declare "$x \neq y$", and terminate.
• If they did not terminate above, they declare "$x = y$".
```

This protocol is very close to our usual randomized protocol for EQ except that now we are choosing $r_1, \ldots, r_n$ conditioned on being linearly independent. This is because the protocol is supposed to be error-free, and thus we have to be able to conclude $X = Y$ if $\langle X, r_i \rangle = \langle Y, r_i \rangle$ for all $i = 1, \ldots, n$.

It is very interesting to see that in the case of $X = Y$ how much information Alice and Bob reveal to each other in the above protocol. Note that in this case, all the information that Alice is sending to Bob (i.e. the n bits $\langle X, r_1 \rangle, \ldots, \langle X, r_n \rangle$) worth 1 bit of information. For example, if instead someone sends 1 bit to Bob in the beginning indicating that $X = Y$, then Bob could recover all these n bits from that 1 bit. On the other hand if $X \neq Y$, then every new bit that Bob sends to Alice can worth almost 1 bit of new information about X for Bob. However, fortunately if $X \neq Y$, the expected number of rounds that it takes for the protocol to find out that $X \neq Y$ and terminate is $O(1)$. Hence, the protocol terminates before Alice and Bob reveal too much information about their inputs. Let us state this formally as a theorem.

Theorem 3. *The above protocol solves equality with no error, and*

$$\mathrm{IC}_\mu(\pi) = O(1),$$

for every distribution μ.

Proof. Note that if $x \neq y$, then since at round j, there will be at most 2 bits of communication, the expected number of communicated bits is

$$\sum_{j=1}^{n} 2\Pr_{r_1,\ldots,r_{j-1}}[\langle x, r_i \rangle = \langle y, r_i \rangle \ \forall i = 1, \ldots, j-1] = 2\sum_{j=1}^{n} \frac{2^{n-j+1}-1}{2^n - 1} \leq 2\sum_{j=1}^{n} \frac{2^{n-j+1}}{2^n} \leq 4.$$

Here $\frac{2^{n-j+1}-1}{2^n-1}$ is the probability that a nonzero $z \in \mathbb{F}_2^n$ (corresponding to $x - y$) belongs to the kernel of the matrix with rows $r_1, \ldots, r_{j-1}$. By symmetry this kernel is uniformly distributed over all $(n-j+1)$-dimensional subspaces of $\mathbb{F}_2^n$.

Let us analyze the information complexity. Consider the random variable $B = 1_{[X=Y]}$. Then

$$\begin{aligned} I(X;\Pi|Y) &= I(XB;\Pi|Y) \leq 1 + I(X;\Pi|YB) = 1 + \mathbb{E}_{b\sim B} I(X;\Pi|Y,(B=b)) \\ &\leq 1 + I(X;\Pi|Y,(X=Y)) + I(X;\Pi|Y,(X\neq Y)) = 1 + I(X;\Pi|Y,(X\neq Y)) \\ &\leq 1 + \mathrm{CC}_{\mu|_{X\neq Y}}(\pi) \leq 5. \end{aligned}$$

Hence $\mathrm{IC}(\pi,\mu) \leq 10$.

□

What is the external communication complexity of the above protocol? Note that in the case where $X = Y$, the external observer learns a lot about X and Y. Since unlike Bob, the external observer does not know the value of Y, each new bit $\langle X, r_i \rangle$ can provide almost a new bit of information about X (for example if μ is uniformly distributed on $\{(x,x) : x \in \{0,1\}^n\}$. Indeed as we shall see below every protocol that solves EQ with zero error on every input must have external information complexity at least n with respect to this measure. This also provides a lower bound for communication cost as it is always greater or equal to the external information cost.

Theorem 4. *Let π be a protocol that computes* EQ *with no error on all inputs $(x,y) \in \{0,1\}^n \times \{0,1\}^n$, and let μ be the measure on on this set that is uniformly distributed on its support $\{(x,x) : x \in \{0,1\}^n\}$. Then*

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

Proof. As we discussed above, we can assume that π does not use public randomness. First note that for every transcript t, there is at most one $a \in \{0,1\}^n$ such that $\Pr[\Pi_{aa} = t] \neq 0$. Otherwise there would be $x \neq y$ with $\Pr[\Pi_{xx} = t] = P^x_A(t)P^x_B(t) > 0$ and $\Pr[\Pi_{yy} = t] = P^y_A(t)P^y_B(t) > 0$, and this would imply that $\Pr[\Pi_{xy} = t] = P^x_A(t)P^y_B(t) > 0$. But this contradicts the assumption that π solves EQ with no error on all inputs, as then t cannot happen with nonzero probability for both xx and xy.

Consequently $H(XY|\Pi) = 0$ as μ is supported on $\{(x,x) : x \in \{0,1\}^n\}$, and by the above observation, the value of x is uniquely determined from the transcript Π. Hence

$$\mathrm{IC}^{\mathrm{ext}}_{\mu}(\pi) = I(XY, \Pi) = H(XY) - H(XY|\Pi) = H(XY) = n.$$

□

References

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