COMP760, SUMMARY OF LECTURE 20.

HAMED HATAMI

1. The internal communication cost of AND (Upper-bound)

The material for today's lecture is from [BGPW13].

Determining the 0-error internal information cost of the AND function is considerably more difficult that determining its external information cost. At the first glance it might seem surprising that one can do something different from the protocol that we discussed in the external case. First we state a protocol that is optimal for the case of symmetric μ, i.e. $\mu(01) = \mu(10)$. The protocol as it is described below is not a conventional protocol. It has access to a clock that continuously increases from 0 to 1.

Figure 1. The optimal protocol π_{sym} for solving AND on a symmetric distribution μ.

- If $x = 0$, Alice samples $\alpha_A \in [0, 1)$ uniformly at random. If $x = 1$ she sets $\alpha_A = 1$.
- If $y = 0$, Bob samples $\alpha_B \in [0, 1)$ uniformly at random. If $y = 1$ he sets $\alpha_B = 1$.
- Alice and Bob monitor the clock, and when the clock reaches $\min(\alpha_A, \alpha_B) < 1$, the corresponding player sends 0 to the other player, and the protocol ends. If the clock reaches $\min(\alpha_A, \alpha_B) = 1$, then they both know that the output is 1.

Note that by replacing the clock with a discrete clock we can approximate the above protocol by a conventional protocol. Indeed we can initially set the time $t = 0$, and then at every round increase it by $\epsilon = \frac{1}{K}$ for a large K. Now at every round Alice and Bob communicate to verify that still $t < \min(\alpha_A, \alpha_B)$. Hence this protocol has communication $O(K)$, and as we tend K to infinity, the information cost of the protocol will converge to $\mathrm{IC}_\mu(\pi_{\mathrm{sym}})$. In fact it is possible to show that no finite-round protocol can achieve the optimal information cost for the AND function and thus it is necessary to take the infimum in the definition of the information cost. Let us analyze the information cost of the above protocol.

Proposition 1. *Let μ be a symmetric measure and*

$$\begin{bmatrix} \mu(00) & \mu(01) \\ \mu(10) & \mu(11) \end{bmatrix} := \begin{bmatrix} \alpha & \beta \\ \beta & \delta \end{bmatrix}.$$

For the above protocol π_{sym}, we have

$$\mathrm{IC}_\mu(\pi_{\mathrm{sym}}) = \frac{\beta}{\ln 2} + 2\delta \log \frac{\delta + \beta}{\delta} + 2\beta \log \frac{\delta + \beta}{\beta} + \frac{\beta^2}{\alpha} \log \frac{\beta}{\alpha + \beta} + \alpha \log \frac{\alpha + \beta}{\alpha}$$

Proof. First note that the protocol is symmetric with respect to X and Y. Hence

$$\mathrm{IC}_\mu(\pi) = 2I(X; \Pi | Y).$$

We have

$$I(X;\Pi|Y=1)=H(X|Y=1)-H(X|\Pi,Y=1)=H(X|Y=1)$$

as given $Y=1$, the transcript Π determines X. Hence

$$\begin{array}{rcl} I(X;\Pi|Y) & = & (\alpha+\beta)I(X;\Pi|Y=0)+(\beta+\delta)I(X;\Pi|Y=1) \\ & = & (\alpha+\beta)I(X;\Pi|Y=0)+(\beta+\delta)H(X|Y=1) \\ & = & (\alpha+\beta)\mathbb{E}_{x\sim X|_{Y=0}}\left[D(\Pi_{XY=x0}\|\Pi_{Y=0})\right]+(\beta+\delta)H(X|Y=1) \\ & = & \alpha D(\Pi_{XY=00}\|\Pi_{Y=0})+\beta D(\Pi_{XY=10}\|\Pi_{Y=0})+(\beta+\delta)H(X|Y=1). \end{array}$$

A transcript Π on x,y can be represented uniquely by the value $t\in[0,1]$ of the clock when the protocol terminated together with a name of the player $\mathcal{P}\in\{A,B\}$ whose random number is reached by the clock first. Hence

$$D(\Pi_{XY=xy}\|\Pi_{Y=y})=\sum_{\mathcal{P}\in\{A,B\}}\int_0^1 f_{xy}(t,\mathcal{P})\log\frac{f_{xy}(t,\mathcal{P})}{f_y(t,\mathcal{P})}dt,$$

where $f_{xy}(t,\mathcal{P})$ and $f_y(t,\mathcal{P})$ are respectively the PDF for $\Pi_{XY=xy}$ and $\Pi_{Y=y}$. We have

- $f_{00}(t,A)=f_{00}(t,B)=1-t$ for $t\in[0,1]$.
- $f_{10}(t,A)=0$ for $t\in[0,1)$ and $f_{10}(t,B)=1$ for $t\in[0,1)$.
- $f_0(t,A)=\frac{\alpha}{\alpha+\beta}(1-t)$ for $t\in[0,1]$ and $f_0(t,B)=\frac{\beta}{\alpha+\beta}+\frac{\alpha}{\alpha+\beta}(1-t)$ for $t\in[0,1)$.

Using these facts we obtain

$$\begin{array}{rcl} I(X;\Pi|Y) & = & \alpha\displaystyle\int_0^1\left((1-t)\log\frac{\alpha+\beta}{\alpha}+(1-t)\log\frac{(1-t)(\alpha+\beta)}{\beta+(1-t)\alpha}\right)dt \\ & & +\beta\displaystyle\int_0^1\log\frac{\alpha+\beta}{\beta+(1-t)\alpha}dt+(\beta+\delta)H\left(\frac{\beta}{\beta+\delta}\right). \end{array}$$

After simplifying this, we obtain

$$\mathrm{IC}_\mu(\pi)=I(X;\Pi|Y)=\frac{\beta}{\ln 2}+2\delta\log\frac{\beta+\delta}{\delta}+2\beta\log\frac{\beta+\delta}{\beta}+\frac{\beta^2}{\alpha}\log\frac{\beta}{\alpha+\beta}+\alpha\log\frac{\alpha+\beta}{\alpha}.$$

□

Now we turn to the non-symmetric case. In this case the trick is that first one of the players sends a bit so that conditioned on this bit, the distribution becomes symmetric. Let μ be the measure

$$\begin{bmatrix}\mu(00) & \mu(01)\\ \mu(10) & \mu(11)\end{bmatrix}:=\begin{bmatrix}\alpha & \beta\\ \gamma & \delta\end{bmatrix}.$$

Let us see what happens in the non-symmetric case. Suppose without loss of generality that $\beta<\gamma$. Then Bob sends the bit 1 with probability $\rho:=\beta+\delta+(\alpha+\gamma)\beta/\gamma=\delta+2\beta+\alpha\beta/\gamma$. If Bob sends 0, the protocol terminates, otherwise the players continue running the previous protocol (for the symmetric case) on the *symmetric* distribution

$$\tilde{\mu}:=\mu|_{B=1}=\begin{bmatrix}\tilde{\mu}(00) & \tilde{\mu}(01)\\ \tilde{\mu}(10) & \tilde{\mu}(11)\end{bmatrix}:=\begin{bmatrix}\frac{\alpha\beta}{\gamma\rho} & \frac{\beta}{\rho}\\ \frac{\beta}{\rho} & \frac{\delta}{\rho}\end{bmatrix}.$$

Hence the information cost of the protocol for the case $\beta\le\gamma$ is given by

$$\mathrm{IC}_\mu(\pi_\wedge)=I(Y;B|X)+\rho\,\mathrm{IC}_{\tilde{\mu}}(\pi_\wedge). \qquad (1)$$

FIGURE 2. The optimal protocol $\pi_\wedge$ for solving AND on an arbitrary distribution μ.

- If $\beta < \gamma$ (resp. $\beta > \gamma$), then Bob (resp. Alice) sends a random bit B as follows
$$B := \begin{cases} 1 & y = 1 \\ 0 & \text{with probability } 1 - c \text{ if } y = 0 \\ 1 & \text{with probability } c \text{ if } y = 0 \end{cases}$$
where $c = \beta/\gamma$ (resp. $c = \gamma/\beta$).
- If $B = 0$ the ptotocol terminates and the players output 0, otherwise run the protocol π_{sym}:
  - If $x = 0$, Alice samples $\alpha_A \in [0, 1)$ uniformly at random. If $x = 1$ she sets $\alpha_A = 1$.
  - If $y = 0$, Bob samples $\alpha_B \in [0, 1)$ uniformly at random. If $y = 1$ he sets $\alpha_B = 1$.
  - Alice and Bob monitor the clock, and when the clock reaches $\min(\alpha_A, \alpha_B) < 1$, the corresponding player sends 0 to the other player, and the protocol ends. If the clock reaches $\min(\alpha_A, \alpha_B) = 1$, then they both know that the output is 1.

Note that we have already calculated the value of $\mathrm{IC}_{\tilde{\mu}}(\pi_\wedge)$ in Proposition 1. Finally since the role of Alice and Bob is symmetric, the case $\beta \geq \gamma$ follows immediately
$$\mathrm{IC}_\mu(\pi_\wedge) = \mathrm{IC}_{\mu^T}(\pi_\wedge).$$
This finishes the analysis of the upper-bound for the information cost of the AND function.

Using numerical analysis, it turns out the hardest distribution for this protocol is given by
$$\mu = \begin{bmatrix} 0.0808931\ldots & 0.264381\ldots \\ 0.264381\ldots & 0.390346\ldots \end{bmatrix}$$
and for this μ, the information cost of above protocol is at most $1.49238\ldots$. Hence we obtain *upper-bound* in the following theorem.

Theorem 2. *We have*
$$\mathrm{IC}(\mathrm{AND}) = 1.49238\ldots$$

The distributions that assign a 0 probability mass to the point 11 are important for the application to the set disjointness problem. The upper-bound in the following theorem easily follows from Proposition 1 and (1).

Theorem 3. *We have*
$$\inf_{\pi} \max_{\mu:\mu(11)=0} \mathrm{IC}_\mu(\pi) = 0.482702\ldots,$$
where the infimum is over all protocols π that correctly compute the AND function on all *inputs (including* 11*).*

In the ext section we finish the proofs of Theorems 2 and 3 by showing that the protocol $\pi_\wedge$ is optimal for every measure μ.

2. Proving lower-bounds for information cost

In this section we will show a method for proving lower-bounds on the information cost of a function. To this end, we would like to be able to consider only protocols that have certain desirable properties. First of all note that we can assume that at every round a player sends only one bit. However now we do not assume that Alice and Bob alternate.

Definition 4. *We say that a protocol π is in normal form if for each fixing r of the public randomness, and each internal node u in the protocol π_r,*

$$\Pr_{xy,R_A,r_B}[\textit{owner of } u \textit{ sends } 0 \mid u] = \frac{1}{2}.$$

Note that the above definition means that the external's estimated distribution q_u^{ext} at every node u is the uniform distribution over its two children. The next lemma shows that one can convert a protocol to normal form adding only an arbitrarily small amount of error without increasing its information cost.

Lemma 5. *Let π be a protocol on inputs $\mathcal{X} \times \mathcal{Y}$ and let μ be a distribution on $\mathcal{X} \times \mathcal{Y}$. For every $\delta > 0$ there exists a protocol π_δ in normal form such that*

1. *π_δ simulates π with error at most δ.*
2. *$\mathrm{IC}_\mu(\pi_\delta) \leq \mathrm{IC}_\mu(\pi)$.*

Proof. Let ℓ be large enough so that $\mathrm{CC}(\pi)/\delta \leq 2^\ell$. In π_δ Alice and Bob will try to simulate the execution of π on (x, y). Consider a node u owned by Alice, and recall that

$$p_u^x(w) = \Pr[\text{Alice sends } w | x, u],$$

$$q_u^{\text{ext}}(w) = \Pr[\text{Alice sends } w | u].$$

Now Alice does the following

- She first chooses a bit B according to her distribution p_u^x, and instead of sending B does the following.
- If $B = 0$ she chooses $\alpha \in [0, q_u^{\text{ext}}(0)]$ uniformly at random, and sends the first ℓ bits of its binary expansion to Bob.
- If $B = 1$ she chooses $\alpha \in (q_u^{\text{ext}}(0), 1]$ uniformly, and sends the first ℓ bits of its binary expansion to Bob.

First note that since Bob knows q_u^{ext}, he will know the value of B if the first ℓ digits of α do not match the first ℓ digits of q_u^{ext}. If they match, Bob declares "Fail" and they terminate. Let's see what is the probability that Bob declares "Fail". Since $q_u^{\text{ext}}(w) = \Pr[B = w|u]$, over a random input xy, with probability $q_u^{\text{ext}}(0)$ the bit B is equal to 0 and with probability $q_u^{\text{ext}}(1) = 1 - q_u^{\text{ext}}(0)$ it is equal to 1. This shows that over a random input xy, the variable α is uniformly distributed in the interval $[0, 1]$. Since q_u^{ext} does not depend on the input, the probability that Bob is not able to recover B is $2^{-\ell}$. Taking the union bound we get that the simulation fails with probability at most $\mathrm{CC}(\pi)2^{-\ell} \leq \delta$.

To show that the information cost does not increase, we focus on this node u. Let B_ℓ denote the vector of the ℓ bits sent by Alice. Then since α determines the value of B_ℓ, we have

$$I(B_\ell; X|Y) \leq I(B\alpha; X|Y) = I(B; X|Y) + I(\alpha; X|YB) = I(B; X|Y),$$

where $I(\alpha; X|YB) = 0$ since conditioned on B, α is a uniform random number in $[0, q_u^{\text{ext}}(0)]$ or $(q_u^{\text{ext}}(0), 1]$, and thus is independent of X. This shows that at the node u, the amount of information revealed about X in π_δ is at most the amount of information revealed in π. Applying this to every node we conclude that

$$\mathrm{IC}_\mu(\pi_\delta) \leq \mathrm{IC}_\mu(\pi).$$

□

Let $\Delta(\mathcal{X}\times\mathcal{Y})$ denote the set of all probability measures on $\mathcal{X}\times\mathcal{Y}$. Given a function f on $\mathcal{X}\times\mathcal{Y}$, we want to introduce a collection $\mathcal{C}(f)$ of functions $C:\Delta(\mathcal{X}\times\mathcal{Y})\to\mathbb{R}^+$ such that

$$\mathrm{IC}_\mu(f)=\max_{C\in\mathcal{C}(f)} C(\mu).$$

Definition 6. *For $f:\mathcal{X}\times\mathcal{Y}\to\mathcal{Z}$ define the family $\mathcal{C}(f)$ of functions $C:\Delta(\mathcal{X}\times\mathcal{Y})\to\mathbb{R}^+$ as all functions satisfying:*

- *$C(\mu)\le\log(|\mathcal{X}|\cdot|\mathcal{Y}|)$ for all μ.*
- *If $f|_{\mathrm{Supp}(\mu)}$ is constant then $C(\mu)=0$.*
- *For all $\mu,\mu_0^A,\mu_1^A\in\Delta(\mathcal{X}\times\mathcal{Y})$ if there is a signal B in normal form that Alice can send starting from μ such that $\mu_j^A(x,y)=\Pr[XY=xy|B=j]$ for $j=0,1$, then*
$$C(\mu)\le\frac{C(\mu_0^A)}{2}+\frac{C(\mu_1^A)}{2}+I(X;B|Y).$$
- *Similarly for Bob*
$$C(\mu)\le\frac{C(\mu_0^B)}{2}+\frac{C(\mu_1^B)}{2}+I(Y;B|X),$$
where $\mu,\mu_0^B,\mu_1^B\in\Delta(\mathcal{X}\times\mathcal{Y})$ and B are defined analogous to the Alice's case.

First note that if π is a 0-error protocol in *normal form* that computes a function f, then a simple induction shows that for every $C\in\mathcal{C}(f)$ we have

$$C(\mu)\le\mathrm{IC}_\mu(\pi).$$

In fact the induction shows more.

Lemma 7. *Let $f:\mathcal{X}\times\mathcal{Y}\to\mathcal{Z}$ be a given function and π be a a protocol in normal form. Then for all $C\in\mathcal{C}(f)$ and all $\mu\in\Delta(\mathcal{X}\times\mathcal{Y})$ we have*

$$C(\mu)\le\mathrm{IC}_\mu(\pi)+\mathbb{E}_{t\sim\Pi}\left[C(\mu_t)\right],$$

where $\mu_t(x,y)=\Pr[XY=xy|t]$.

Proof. Exercise. □

One can combine Lemma 7 and Lemma 5 to obtain the following lemma.

Lemma 8. *Let $f:\mathcal{X}\times\mathcal{Y}\to\mathcal{Z}$ be a given function and τ be a a protocol that solves f correctly on all inputs. Then for all $C\in\mathcal{C}(f)$ and all $\mu\in\Delta(\mathcal{X}\times\mathcal{Y})$ we have*

$$C(\mu)\le\mathrm{IC}_\mu(\tau).$$

Proof. Exercise. □

Finally the following corollary shows that $\mathcal{C}(f)$ provides the exact value of the information complexity of a function.

Corollary 9. *For every $f:\mathcal{X}\times\mathcal{Y}\to\mathcal{Z}$, and every $\mu\in\Delta(\mathcal{X}\times\mathcal{Y})$ we have*

$$\mathrm{IC}_\mu(f)\in\mathcal{C}(f),$$

and

$$\mathrm{IC}_\mu(f)=\max_{C\in\mathcal{C}(f)} C(\mu).$$

2.1. **Protocol as a random walk on distributions.**

Consider a protocol π and a prior distribution μ. Suppose that in the first round Alice randomly sends a message $\Pi_1 \in \mathcal{M}$. Let $\mu_M = \mu|_{\Pi_1=M}$ for $M \in \mathcal{M}$. In other words

$$\mu_M(x,y) := \Pr[XY = xy|\Pi_1 = M].$$

It is not difficult to see that $\mu_M(x,y) = p_M(x)\mu(x,y)$ for some function $p_M(x)$. That is to say μ_M is obtained by multiplying the rows of μ by certain numbers. Similarly if Bob is sending a message then the columns of μ will be multiplied $\mu_M(x,y) = \mu(x,y)p_M(y)$.

Here we can think that Alice is randomly updating the prior distribution μ to a new distribution μ_M. Now consider a different protocol τ where now Alice is randomly sending a message $T_1 \in \mathcal{M}$. Suppose that μ_{Π_1} (this is a random element of $\Delta(\mathcal{X} \times \mathcal{Y})$) have the same distribution as μ_{T_1}. That is for every $\nu \in \Delta(\mathcal{X} \times \mathcal{Y})$ we have

$$\Pr[\mu_{\Pi_1} = \nu] = \Pr[\mu_{T_1} = \nu].$$

In other words, π and τ might perform completely different things, and nevertheless the updated distributions have the same distributions [1]. Then what can we say about the amount of the information leaked in these two protocols? Can the fact that the first step is performed differently change the information leak? The answer turns out to be "No". So for example if in π Alice sends one bit and with probability 1/2 gets to the updated distribution μ_0, and with probability 1/2 to the updated distribution μ_1, and in τ Alice sends one of the possibly many messages and yet always arrives at μ_0 or μ_1, and with equal probability, then the information leakage in both protocols is the same.

Lemma 10. *If in both protocols π and τ, Bob speaks first and the distribution of the updated distribution is the same for both protocols, then*

$$I(X;\Pi_1|Y) = I(X;T_1|Y).$$

The proof is fairly straightforward and we leave it as an exercise.

2.2. **Applying to the AND function.**

The protocol $\pi_\wedge$ discussed for the AND function suggests dividing the set of all measures

$$\mu = \begin{bmatrix} \mu(00) & \mu(01) \\ \mu(10) & \mu(11) \end{bmatrix} := \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix}.$$

into three regions.

- **Bob's region:** all distributions μ with $\beta < \gamma$.
- **Alice's region:** all distributions μ with $\beta > \gamma$.
- **Diagonal region:** all distributions μ with $\beta = \gamma$.

If μ is in Bob's region then in the protocol $\pi_\wedge$ Bob speaks first, and if it is in Alice's region then Alice speaks first. If it is in the diagonal, then they run the symmetric part of the protocol. Note that in $\pi_\wedge$ it never happens that a measure in Bob's region gets updated to a measure in Alice's region or vice versa. In other words to cross from Alice's region to Bob's region (or vice versa), first the measure must go into the diagonal region. We call the signals B that satisfy this property *non-crossing signals.*

Using Lemma 10 it is possible to show that if τ is a protocol that computes the AND function, then one can modify it so that all its signals will become non-crossing. Roughly speaking we replace a crossing signal that updates μ to μ_0 and μ_1 with a random walk that to go from Alice's region

[1] Note that here we are talking about distributions on distributions $\Delta(\mathcal{X} \times \mathcal{Y})$.

to Bob's region always first goes to the diagonal region. The random walk will terminate when it reaches μ_0 or μ_1. We refer the reader to the original paper [BGPW13, Claim 4.17] for a proof.

Now suppose that we have a prior distribution μ in Bob's region. We can start running our protocol $\pi_\wedge$ on μ and that will have cost $\mathrm{IC}_\mu(\pi_\wedge)$. But now suppose that instead Bob starts and sends a non-crossing signal B in normal form, and after that we run the protocol $\pi_\wedge$ (on the updated prior distribution). Note that after sending B, the prior will be updated to μ_0 with probability 1/2 (if $B = 0$) and to μ_1 with probability 1/2 (if $B = 1$). Since B is non-crossing neither μ_0 nor μ_1 is in Alice's region. Let us call this second protocol τ. The following lemma shows that $\mathrm{IC}_\mu(\tau) = \mathrm{IC}_\mu(\pi_\wedge)$. Note that in $\pi_\wedge$ Bob sends a bit so that μ gets updated to a distribution in the diagonal region as quickly as possible. However it follows from the following lemma and the remark following it that Bob can instead meander in his region awhile before reaching the diagonal without any extra information cost.

Lemma 11. *Suppose that the prior μ is in Bob's region and Bob sends a non-crossing signal B in normal form, updating μ to μ_0 or μ_1 with probability 1/2, and then runs $\pi_\wedge$ on the updated distribution. Then*

$$\mathrm{IC}_\mu(\pi_\wedge) = \mathrm{IC}_\mu(\tau) = \frac{\mathrm{IC}_{\mu_0}(\pi_\wedge)}{2} + \frac{\mathrm{IC}_{\mu_0}(\pi_\wedge)}{2} + I(Y;B|X).$$

Proof. See [BGPW13, Lemma 7.15 and Claim 7.16]. □

Remark 12. Note that by repeatedly applying the above lemma we can show that if the prior μ is in Bob's region and if τ is a protocol in which first Bob sends a sequence of signals so that none of the updated distributions are in Alice's region, and afterwards runs $\pi_\wedge$ with the updated distribution, then

$$\mathrm{IC}_\mu(\pi_\wedge) = \mathrm{IC}_\mu(\tau).$$

■

Finally we are ready to show that π is an optimal protocol for the internal cost of the AND function. To this end we need to show that $C_\pi : \mu \mapsto \mathrm{IC}_\mu(\pi)$ belongs to $\mathcal{C}(\mathrm{AND})$ in Definition 6. The only non-trivial condition to check is the one regarding the signal B sent by Alice or Bob. The above discussion shows that one can simulate a protocol with one that uses only non-crossing signals. Consequently, we can assume that B is a non-crossing signal. Suppose that

$$\mu := \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix}.$$

and that Bob sends a non-crossing signal B and μ gets updated to either μ_0 or μ_1 each with probability 1/2. Without loss of generality we can assume that $\Pr[B = 0|Y = 0] \geq \frac{1}{2}$ as $\Pr[B = 0|Y = 0] + \Pr[B = 1|Y = 0] = 1$ and we can switch the role of $B = 0$ and $B = 1$ otherwise. So we assume:

- $\Pr[B = 0] = \Pr[B = 1] = \frac{1}{2}$.
- $\Pr[B = 0|Y = 0] = \frac{1}{2} + \epsilon_0/2$ for $\epsilon_0 \geq 0$.
- $\Pr[B = 1|Y = 1] = \frac{1}{2} + \epsilon_1/2$ for $\epsilon_1 = \epsilon_0\frac{\alpha+\gamma}{\delta} \geq 0$.

Moreover we can assume that $\epsilon_0, \epsilon_1 \geq 0$ are sufficiently small by possibly replacing B with many slightly biased (towards the value B) bits. Then

$$\mu_0 := \begin{bmatrix} (1+\epsilon_0)\alpha & (1-\epsilon_1)\beta \\ (1+\epsilon_0)\gamma & (1-\epsilon_1)\delta \end{bmatrix}.$$

and

$$\mu_1 := \begin{bmatrix} (1-\epsilon_0)\alpha & (1+\epsilon_1)\beta \\ (1-\epsilon_0)\gamma & (1+\epsilon_1)\delta \end{bmatrix}.$$

To finish the proof we need to show that $C_\wedge(\mu) := \mathrm{IC}_\mu(\pi)$ satisfies

$$C_\wedge(\mu) \le \frac{C_\wedge(\mu_0)}{2} + \frac{C_\wedge(\mu_1)}{2} + I(B;Y|X). \tag{2}$$

We consider three cases. The computations are done using Wolfram Mathematica.

- $\mu \in$ **Bob's region:** Then Lemma 11 verifies (2).
- $\mu \in$ **Alice's region:** Then $\beta > \gamma$ and thus
$$\frac{C_\wedge(\mu_0)}{2} + \frac{C_\wedge(\mu_1)}{2} + I(B;Y|X) - C_\wedge(\mu) = \frac{\alpha(\beta-\gamma)}{(\alpha+\beta)\delta^2 \ln 4}\epsilon_0^2 \pm O(\epsilon_0^3) > 0$$
for sufficiently small ϵ_0.
- $\mu \in$ **the diagonal region:** Then $\beta = \gamma$ and thus
$$\frac{C_\wedge(\mu_0)}{2} + \frac{C_\wedge(\mu_1)}{2} + I(B;Y|X) - C_\wedge(\mu) = \frac{\alpha\beta}{12(\alpha+\beta)\delta^3 \ln 2}\epsilon_0^3 \pm O(\epsilon_0^4) > 0$$
for sufficiently small ϵ_0.

This verifies (2) and shows that $C_\wedge(\mu) = \mathrm{IC}_\mu(\pi)$ is a lower-bound for the information cost of the AND function.

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

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