Special Topics in Complexity Theory, Fall 2017. Instructor: Emanuele Viola

1 Separating information and communication, Scribe: Biswaroop Maiti

This is a presentation by Willy Quach regarding separating information complexity and communication complexity [GKR16]. This paper shows an exponential gap between the distributional communication complexity and information complexity. BR11 showed that information complexity and amortized communication complexity are equal. Therefore, this also shows an exponential gap between distributional and amortized communication complexity. This is an extension of [GKR14], where an exponential gap was shown for a relation with large outputs, whereas, here, the separation is shown using a partial boolean function with 1 bit output.

The fundamental question answered here is described as follows. There are two parties Alice(A) and Bob(B), who have inputs $x \in \{0,1\}^n$, and $y \in \{0,1\}^n$, respectively and they want to compute the function $f(x,y)$. We know that if either of them can communicate their input to the other, then the receiving party can compute f and output with one bit. This would require at most $n+1$ bits of communication. However, this communication need not be one sided, both the parties can interact in the process where one party is trying to know the input of the other. If either of them decide to merely transmit their input to the other, this would require exactly n bits. However, the number of bits to compute $f(x,y)$ need not always require the same number of bits. In contrast, the information content of an input $X \in \{0,1\}^n$ distributed according to a given distribution will be given by its Shannon entropy. One could ask whether, in an interactive setting, when A is trying to communicate her input X to B, will she need to communicate $H(X)$ bits?

In this model, we consider the case where there are two parties Alice(A) and Bob(B), who have inputs x, and y, drawn from a joint distribution μ, publicly known and they want to compute the function $f(x,y)$. A and B may have both public and private randomness, though we may assume that they have only private randomness, without loss of generality in our consideration. The inputs, x, y may be correlated. We would define a protocol to compute $f(x,y)$, if it computes f correctly with probability $2/3$ over the randomness of μ, the public random coins and private random coins of A and B. The communication complexity will be defined as the length of the transcript Π.

$$\mathsf{CC}(\pi) = |\Pi|$$

We will sketch some details of the separation between information complexity and communication complexity in the distributional model. In particular, we will show the upper bound on information cost for an explicit partial function called the *bursting noise function*. For this, we will use the notion of *divergence cost*, formally defined in [BR14], that was also implicit in [BBCR13].

1.1 Definitions

Definition 1. We define Shannon entropy of a discrete variable X that takes values $x \in \Omega$ according to a probability distribution μ as

$$H(X) = \sum_{x\in\Omega} -\mu(x)\log(\mu(x)).$$

Similarly, conditional entropy for a random variable $X|Y$ (X conditioned on Y), that take values according to a joint distribution $(X,Y) \sim \mu(x,y)$ is defined as

$$H(X|Y) = \epsilon_y[H(X)|y \in Y],$$

where the expectation is defined with respect to the joint distribution of X, Y.

Definition 2. Let $\mu_1, \mu_2 : \Omega \to [0,1]$ be two distributions, where Ω is discrete (but not necessarily finite). The relative entropy between μ_1, μ_2, denoted $D(\mu_1||\mu_2)$, is defined as

$$D(\mu_1||\mu_2) = \sum_{x\in\Omega} \mu_1(x)\log\frac{\mu_1(x)}{\mu_1(x)}$$

We can extend the notion of Shannon entropy into the information cost of a protocol as defined below.

Definition 3. The information cost of a protocol π over random inputs (X, Y) that are drawn according to a joint distribution μ, is defined as

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

where Π is a random variable which is the transcript of the protocol π with respect to μ. That is, Π is the concatenation of all the messages exchanged during the execution of π. The information cost of a computational task f with respect to a distribution μ is defined as

$$\mathsf{IC}\mu(f, \epsilon) = \inf_{\pi} \mathsf{IC}\mu(\pi),$$

where the infimum ranges over all protocols π that solve f with error at most ϵ on inputs that are sampled according to μ.

To separate CC and IC, we need to exhibit a function f, computed with respect to a joint distribution μ, such that the lower bound of $\mathsf{CC}_{\mu}(f, \epsilon)$ and the upper bound on $IC_{\mu}(f)$ is large, in this case exponential in k, an input parameter. In this case, the function is explicit but a partial function.

Claim 4.

$$\mathsf{CC}(f, \mu) > \mathsf{IC}(f, \mu)$$

In order to separate IC and CC, we should show a separation of lower bound on CC and upper bound on IC.

Theorem 5. $\forall (f, \mu)$ such that $\mathsf{IC}(f, \mu) = O(k)$, then there is a protocol π' for computing f such that $\mathsf{CC}(\pi') = 2^{O(k)}$.

Theorem 6. $\forall k$, there exists (f, μ) such that

- There exists a protocol π such that $\mathsf{IC}(\pi) = O(k)$
- For any protocol π, with $\mathsf{CC}(\pi) \leq 2^k$, π fails with probability $(\frac{1}{2} - \varepsilon)$, where $\varepsilon = 2^{-k}$.

$\mathsf{IC}(f, \mu) = O(k)$, then $\exists \pi'$ with $\mathsf{CC}(\pi') = 2^{O(k)}$.

1.2 Bursting Noise Function

We will define the function f and the distribution μ below. Let x and y be the inputs of the players A and B respectively. The *bursting noise function* is a communication game with an integer parameter $k \in \mathbb{N}$, here played between A and B on a directed binary tree $\mathcal{T}$ of $c \cdot w$ layers, $c = 2^{4^k}, w = 2^{100}k$, which are partitioned into c multi-layers , each of width w. For every vertex $v \in V(\mathcal{T})$, there are two labels x_v, y_v, that denote a bit from each of the inputs i.e. $x, y \in \{0,1\}^{|V|}$. Therefore, the length of each input is doubly exponential in k.

The tree has three kinds of multi-layers. There is one multi-layer i that is picked uniformly at random. For all the vertices at layers $< i$, the bits are equal at every vertex i.e. $x_v = y_v$, though the actual value is picked uniformly at random. At the multi-layer i, the input bits are **noisy** i.e. x_v, y_v are bits picked at random. For the multi-layers $> i$, vertices that lead to the good paths, the bits are same, and for vertices on "bad" paths, the input bits are uniformly at random. Alice is said to own the odd layers, i.e. from root onwards, and Bob owns the even layers, i.e. till the leaves.

Now, we precisely define the distribution μ on $\{0,1\}^{|V|} \times \{0,1\}^{|V|}$, where V is the set of vertices in $\mathcal{T}$, as an algorithm.

1. Pick a multi-layer uniformly at random: $i \in_{\$} [c]$ (We will make it noisy).

2. For all $v \in \mathcal{T}$:

   (a) If the depth of $d(v) < i \cdot w$, pick $x_v = y_v \overset{\$}{\leftarrow} \{0,1\}$.

   (b) If in the i-th multi-layer, then pick $x_v \overset{\$}{\leftarrow} \{0,1\}, y_v \overset{\$}{\leftarrow} \{0,1\}$

   (c) For the subtree rooted on v_{xy}, do step (a).

   (d) Else, do, (b).

   (e) Pick a random bit $b \overset{\$}{\leftarrow} \{0,1\}$

   (f) For every leaf: and set $x_v \leftarrow x_v \bigoplus b$.

The goal of the players A, B is to retrieve the random bit b set at the last step.

1.3 Protocol π

The game of *bursting noise function* can proceed in a particular protocol π, which we define now. The two parties A and B are assumed to have no public randomness. Their respective inputs x, y are sampled according to a joint distribution that is known to both the parties. Also, without loss of generality, we assume that the protocol proceeds by the players alternately communicating bit by bit i.e. A sends her first bit; B sends his first bit; then, A sends her second bit, and so on. Therefore, in odd rounds, A sends a bit; in even rounds, B sends a bit. This induces a tree, we denote by $\mathcal{T}_\pi$, associated with the protocol π, in which every vertex v has two outgoing directed edges to its children, labeled 0 and 1. If the v is in the odd layer, the label corresponds to the situation when A sends the corresponding bit, or we say A owns the layer. We assume a convention, that the left edge is labeled by 0, and the right edge is labeled 1. The even layers correspond to B, likewise.

Consider $\ell = 2^{100k}$. At any level, when ℓ or more bits have been transmitted by A, the second player B looks back at the last ℓ of the vertices corresponding to odd layers and checks if those of A's inputs aka labels in the path followed match with his own input bits. If and only if less than $.8\ell$ of those match, then B aborts the protocol.

For every vertex v, we define two sets of distributions P_v and Q_v. We define P_v at first. Recall, every vertex in $\mathcal{T}$ has labels x_v, y_v. For a non-leaf vertex in an odd layer, if $x_v = 0$, the A sends 0 with probability 0.9, and 1 with probability 0.1. We write it succinctly as $P_v = (0.9, 0.1)$. Likewise, if $x_v = 1$, the A sends 1 with probability 0.9, and 0 with probability 0.1, i.e. $P_v = (0.1, 0.9)$. Similarly, for B. If v is in the even layer, if $y_v = 0$, then $P_v = (0.9, 0.1)$ for B. So, associated with $\mathcal{T}_\pi$, for every vertex v, there is $P_v = (p_v, 1 - p_v)$ associated, which says that the next bit to be transmitted is 0, with probability p_v.

Q_v is the estimation of P_v by the opposite player i.e. if v is in the odd layer, then the estimation is by B; otherwise, A. Therefore, associated with the tree $\mathcal{T}$, is the protocol tree $\mathcal{T}_\pi$, and they come with the distributions $\{P_v\}_{v \in V}, \{Q_v\}_{v \in V}$.

We will state without proof, the following lemma.

Lemma 7. Let $(x, y) \in \mathsf{supp}(\mu)$ be an input pair for the bursting noise function. The protocol above π *aborts* with probability at most 2^{-10k} on the input (x, y).

Now, we define the notion of Divergence cost, with which we will define the divergence cost of $\mathcal{T}_\pi$, with respect to Q_v, denoted by $D(\mathcal{T}_\pi^{Q_v})$.

Definition 8. Divergence Cost: Consider a binary tree $\mathcal{T}$, whose root is r, and distributions $P_v = (p_v, 1 - p_v)$, $Q_v = (q_v, 1 - q_v)$ for every non-leaf vertex v in the tree. We think of P_v and Q_v as distributions over the two children of the vertex v. We define the *divergence cost* of the tree T recursively, as follows.

$$D(v \in \mathcal{T}) = \begin{cases} 0 & \text{if } v \text{ is a leaf} \\ D(\mathcal{T}) = D(P_r||Q_r) + \epsilon_{v \sim P_r}[D(\mathcal{T}_v)] & \text{otherwise,} \end{cases}$$

where for every vertex v, $\mathcal{T}_v$ is the subtree of $\mathcal{T}$ whose root is v.

An equivalent definition of the divergence cost of $\mathcal{T}$ is obtained as

$$D(\mathcal{T}) = \sum_{v \in V} \tilde{p_v} \cdot D(P_v||Q_v)$$

where V is the vertex set of $\mathcal{T}$, and for a vertex $v \in V$, $\tilde{p_v}$ is the probability to reach v by following the distributions P_v, starting from the root. Formally, if v is the root of the tree $\mathcal{T}$, then $\tilde{p_v} = 1$, otherwise

$$\tilde{p_v} = \begin{cases} \tilde{p_u} \cdot p_u & \text{if } v \text{ is the left-hand child of } u \\ \tilde{p_u} \cdot (1 - p_u) & \text{if } v \text{ is the right-hand child of } u \end{cases}$$

Set $Q_v^* := \epsilon_x[P_v|y, \pi_v]$.

Lemma 9. For every protocol π and distributions Q_v known to a player who doesn't own v, as above, it holds that:

$$\epsilon(D(\mathcal{T}_\pi^{Q_v^*})) \leq \epsilon(D(\mathcal{T}^{Q_v}\pi))$$

Now, we are ready to sketch our final claim.

Theorem 10. π has information cost $O(k)$. In other words,

$$D(\mathcal{T}_\pi) = O(k).$$

Proof. Fix $(x, y) \in \mathsf{supp}(\mu_i)$ for some $i \in [c]$. From previous lemma, $D(\mathcal{T}_\pi) = \sum_v \tilde{p_v} \cdot D(P_v||Q_v)$.

We need to bound the sum in the last expression separately, for odd and even layers. We will leave it as an exercise to show that the cost for the even layers will be $O(w)$. We will skip the part for the even layers, and sketch the part about the odd layers. A vertex in $\mathcal{T}_\pi$ corresponds to a vertex in $\mathcal{T}$. Therefore, we look at the corresponding structure in $\mathcal{T}$.

Let us consider a vertex $v \in \mathcal{T}_\pi$. If v corresponds to a non-noisy vertex in $\mathcal{T}_\pi$, then we must have $D(P_v||Q_v) = 0$. On the other hand, if v corresponds to a vertex in the noisy multilayer i, then, $D(P_v||Q_v) \leq 4$. Now consider the sum in the above expression of the divergence cost. The vertices above the noisy multi-layer i add zero divergence cost. The vertices of the noisy multi-layer together add $O(w)$ divergence cost. At the end, we are left with the vertices that are below the noisy multi-layer.

Now, let us assume $i < c$, which implies that there is at least one multilayer of vertices below the noisy multi-layer. Let v be in the layer $i^* + w$ of the tree $\mathcal{T}$. Now, if v is a typical vertex, i.e. it is on a good path, it does not add to the divergence cost sum.

Therefore, the only case left is when v is in a layer below the noisy multilayer and is a non-typical vertex. As stated above **state above**, on the non-typical vertices the protocol aborts after 4ℓ rounds, in expectation. The probability that v is a non-typical vertex with respect to multi-layer i, is at most 2^{-1000k}. Therefore, the divergence cost added by this case is $\leq 2^{-1000k} \cdot 4\ell \cdot 4 \leq 1$.

Finally, considering all these cases, we see that the cost for the even vertices is $O(k) = O(w)$. □

By the above claim, we can state the upper bound result as:

Theorem 11. There exists a randomized protocol $\pi$for the *bursting noise function* with parameter k, that errs with probability $\leq 2^{-k}$ (over the input distribution μ) which has

$$\mathsf{IC}_\mu(\pi) = O(k)$$

1.4 Lower Bound

The lower bound on the Communication Complexity is given by lower bound on the distributional communication complexity of the *bursting noise* function using the relative discrepancy method defined as below.

Definition 12. Let $\epsilon \in (0, 1/2)$ and $\delta \in (0, 1)$. Let μ be a distribution over $\{0,1\}^n \times \{0,1\}^n$ and let $f \colon \mathsf{supp}(\mu) \to \{0,1\}$ be a function. We say that (f, μ) has the (ϵ, δ) relative discrepancy property if there exists a distribution ρ over $\{0,1\}^n \times \{0,1\}^n$ such that for every rectangle $R = A \times B \subseteq \{0,1\}^n \times \{0,1\}^n$ with $\rho(R) \geq \delta$, the following two properties hold:

- $\mu(R \cup f^{-1}(0)) \geq (\frac{1}{2} - \epsilon) \cdot \rho(R)$,

- $\mu(R \cup f^{-1}(1)) \geq (\frac{1}{2} - \epsilon) \cdot \rho(R)$.

References

[BBCR13] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. *SIAM Journal on Computing*, 42(3):1327–1363, 2013.

[BR14] Mark Braverman and Anup Rao. Information equals amortized communication. *IEEE Transactions on Information Theory*, 60(10):6058–6069, 2014.

[GKR14] Anat Ganor, Gillat Kol, and Ran Raz. Exponential separation of information and communication. In *Foundations of Computer Science (FOCS), 2014 IEEE 55th Annual Symposium on*, pages 176–185. IEEE, 2014.

[GKR16] Anat Ganor, Gillat Kol, and Ran Raz. Exponential separation of information and communication for boolean functions. *Journal of the ACM (JACM)*, 63(5):46, 2016.