COMP760, SUMMARY OF LECTURE 16.

HAMED HATAMI

1. Direct product and direct sum theorems

One class of the most interesting and fundamental questions of theoretical computer science are the closely related direct product and direct sum questions. A *direct product theorem* in a particular computational model asserts that the probability of success of performing n independent copies of a computational task decreases in n. For example, a direct product theorem might say that if f is computable with a given amount of resources (e.g. computational, communicational, space, etc) with success probability at least $1-\epsilon$ (e.g. on every input, or on a random input, etc), then given the same amount of resources, one can only compute the function $f^n : (x_1, \ldots, x_n) \mapsto (f(x_1), \ldots, f(x_n))$ with success probability at most δ_n where δ_n is much smaller than $1 - \epsilon$. Note that this is harder than performing n copies of the original task as here we want all the coordinates to be correct at the same time. Two fundamental results in computer science, Yao's XOR Lemma [Yao82] and Raz's Parallel Repetition Theorem [Raz98] are examples of direct product theorems.

Sometimes to stay at the Boolean regime, instead of f^n, one considers the Boolean function $f^{\oplus n} : (x_1, \ldots, x_n) \mapsto f(x_1) \oplus \ldots \oplus f(x_n)$. It is in general expected that this function is almost as difficult as f^n. Indeed to have a good estimate for the parity $f(x_1) \oplus \ldots \oplus f(x_n)$ one has to have good estimates for all the values $f(x_1), \ldots, f(x_n)$ as changing any one of them will completely change the value of the parity. However one has to be slightly careful as every Boolean function can be computed with success probability $1/2$ in an obvious manner by just flipping a random coin for each input. Hence in this context, if the original function f can be computed with success rate $\frac{1}{2} + \epsilon$ with the given resources, a direct product theorem would state that $f^{\oplus n}$ can be computed with the same amount of resources only with success probability at most $\frac{1}{2} + \epsilon_n$ where ϵ_n is much smaller than ϵ.

On the other hand the direct sum question fixes the success probability, and asks how much resources one does need to solve f^n compared to the amount of resources needed to solve f (with the same success probability). In other words, a *direct sum theorem* asserts that the amount of resources needed to perform n independent copies of the same task grows with n. While the direct sum question for general models such as Boolean circuits has a long history, no general results are known, and indeed they cannot be achieved by the standard reductions used in complexity theory. Indeed if we start by assuming that there is a circuit C that performs n independent copies of the task, then it is impossible to use this C as a blackbox to show that there is a much smaller circuit C' that solves one copy of the task. Including C as a blackbox will already make the size of the new circuit larger. Indeed it is known that at least the most straightforward and optimistic formulation of a direct sum theorem for Boolean circuits is false. The example comes from fast matrix multiplication. By a counting argument, there exists an $n \times n$ matrix A over GF(2) such that the map $x \mapsto Ax$ requires a circuit of $\Omega(n^2/\log n)$ size. But the map $(x_1, \ldots, x_n) \mapsto (Ax_1, \ldots, Ax_n)$ is just the product of the two matrices A and X (whose columns are $x_1, \ldots, x_n$) and hence can be carried out by a circuit of size $O(n^{2.38}) \ll n^2/\log n$. Unlike in circuit complexity and communication complexity, in information complexity one can use a protocol as a blackbox and still be able to obtain a protocol with a much smaller information cost. We will this in the proof of Theorem 1 below. Indeed this great advantage will be used in many applications of information complexity to communication complexity.

In the context of communication complexity, the direct sum question was first raised by Karchmer, Raz, and Wigderson [KRW95] who conjectured a certain direct sum statement for a certain deterministic communication problem. They showed that this conjectured direct sum statement would separate the two complexity classes NC^1 and NC^2, and in particular show that $\mathrm{NC}^1 \neq \mathrm{P}$. Recall that NC^i is the class of problems that can be solved by a uniform family of Boolean circuits, with polynomial size, depth $O(\log^i(n))$, and fan-in 2. This would be an incredible result as $\log(n)$ seems to be a serious barrier to almost all known techniques for proving general circuit lower-bounds. We will discuss this problem later in the course when we talk about the applications of communication complexity to circuit complexity.

Feder, Kushilevitz, Naor, and Nisan [FKNN95] gave a direct sum theorem for non-deterministic communication complexity, and deduced from it a somewhat weaker result for deterministic communication complexity: if a single copy of a function f requires C bits of communications, then n copies require $\Omega(\sqrt{C}n)$ bits. In other words,

$$\lim_{n\to\infty} \frac{D(f^n)}{n} = \Omega(\sqrt{D(f)}).$$

Feder *et al* also considered the direct sum question for randomized communication complexity and showed that the dependence of the communication on the error of the protocol for many copies can be better than what can be obtained from the naive protocol for many copies. Shaltiel [Sha03] gave a direct sum/product theorem for the discrepancy of a function, which shows a direct sum/product theorem for the randomized communication complexity of the function if discrepancy lower-bound matches the communication complexity of the function. More precisely he showed that $\mathrm{disc}(f^{\oplus n}) = \mathrm{disc}(f)^{\Omega(n)}$. This shows that if a function f is proved to be hard on average for c-bit communication protocols via the "discrepancy method", then $f^{\oplus n}$ is exponentially harder on average for $\Omega(nc)$-bit communication protocols. There are subsequent works on the direct sum question for randomized communication complexity and the distributional communication complexity. Many of these works are based on studying the external information cost. Barak, Braverman, Chen, and Rao [BBCR10] defined the internal information cost, and used it to prove a strong lower-bound on the direct sum problem for the distributional communication complexity. First let us see how information complexity behaves with respect to performing multiple copies of the same task.

Theorem 1 (Additivity of information cost)**.** *For every communication task T, we have*

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

Proof. It is obvious that $\mathrm{IC}_{\mu^n}(T^n) \leq n\mathrm{IC}_\mu(T)$. Indeed given any protocol π for T, we can define a protocol for T^n by running π on each coordinate. Obviously the information cost of this new protocol is $n\mathrm{IC}_\mu(T)$ and it performs T^n.

The opposite directions is more interesting. Consider a protocol π for T^n how can we use π as a blackbox to obtain a protocol that performs T but has information cost $\frac{\mathrm{IC}_\mu(\pi)}{n}$? We define the protocol τ in the following manner.

- On input $(\mathbf{x}, \mathbf{y})$, Alice and Bob
- Publicly choose
  - $i \in \{1, \ldots, n\}$ uniformly at random, and set $X_i = \mathbf{x}$ and $Y_i = \mathbf{y}$ individually.

  - $X_1, \dots, X_{i-1}$ independently, each according to μ_x.
  - $Y_{i+1}, \dots, Y_n$ independently, each according to μ_y.
- Alice privately chooses $X_{i+1}, \dots, X_n$ according to μ conditioned on the values of $Y_{i+1}, \dots, Y_n$.
- Bob privately chooses $Y_1, \dots, Y_{i-1}$ according to μ conditioned on the values of $X_1, \dots, X_{i-1}$.
- They run the protocol π on (X, Y) with $X = (X_1, \dots, X_n)$ and $Y = (Y_1, \dots, Y_n)$, and output the i-th coordinate.

So the above protocol chooses X and Y in the following manner so that the input $(\mathbf{x}, \mathbf{y})$ is embedded in the i-th coordinate and furthermore each (X_i, Y_i) is chosen independently with distribution μ:

$$X = (\underbrace{X_1, \dots, X_{i-1}}_{\text{pub}}, \mathbf{x}, \underbrace{X_{i+1}, \dots, X_n}_{\text{private}}) \qquad \text{and} \qquad Y = (\underbrace{Y_1, \dots, Y_{i-1}}_{\text{private}}, \mathbf{y}, \underbrace{X_{i+1}, \dots, X_n}_{\text{public}}).$$

This protocol τ obviously performs the task T, as it was assumed that π performs T^n. It remains to show that the information cost of τ equals $\mathrm{IC}_{\mu^n}(\pi)/n$.

$$\begin{aligned} I(\mathbf{y}; \Pi_\tau | \mathbf{x}) &= I(\mathbf{y}; \Pi_\pi | i, X, Y_{i+1}, \dots, Y_n) \qquad \text{(conditioning on what Alice knows)} \\ &= \frac{1}{n} \sum_{j=1}^{n} I(Y_i; \Pi_\pi | X, Y_{i+1}, \dots, Y_n, [i = j]) \qquad \text{(expanding the conditioning on } i\text{)} \\ &= \frac{1}{n} \sum_{j=1}^{n} I(Y_j; \Pi_\pi | X, Y_{j+1}, \dots, Y_n) \\ &= \frac{1}{n} I(Y; \Pi_\pi | X) \qquad \text{(chain rule)} \end{aligned}$$

Note that since $X_{i+1}, \dots, X_n$ are determined by Alice's private randomness and i and $X_1, \dots, X_{i-1}$, and $Y_{i+1}, \dots, Y_n$ are determined by public randomness, the first equality basically corresponds to the general identity $I(Y; \Pi | X) = I(Y; \Pi | X, R_A, R)$ which follows from what we saw on lecture 14.

It follows that

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

□

Note that although τ uses π as a blockbox, it nevertheless has smaller information cost. Contrast this with communication protocols, or circuits.

In the first glance the way that the chain rule worked out in the above proof might look rather magical, but indeed the way we constructed X and Y are based on "reversed engineering" of the chain rule

$$\sum_{j=1}^{n} I(Y_j; \Pi_\pi | X, Y_{j+1}, \dots, Y_n) = I(Y; \Pi_\pi | X).$$

2. Compression

Given a task T and a protocol π for T^n, Theorem 1 allows one to obtain a protocol for T with the much smaller information cost of $\mathrm{IC}_\mu(\pi)/n$. Note however that this protocol's communication complexity equals to that of π and thus can be very large. Hence it is desirable to use the small information cost to compress this protocol to a new protocol with small communication cost. Indeed compression is one of the key concepts that allows one to apply information complexity to communication complexity. The following table summarizes the current known results about compression.

Table 1. Compression results: $I = \mathrm{IC}_\mu(\pi)$, $I^{\mathrm{ext}} = \mathrm{IC}_\mu^{\mathrm{ext}}(\pi)$, $C = \mathrm{CC}(\pi)$, and k is the number of rounds.

reference	worst case CC of the compressed protocol in $O(\cdot)$	in $O_\epsilon(\cdot)$
[BBCR10]	$\sqrt{IC}\log(C/\epsilon)/\epsilon$	$\sqrt{IC}\log(C)$
[BR14]	$\frac{1}{\epsilon} \times (I + k\log(k) + \sqrt{kI})$	$I + k\log(k) + \sqrt{kI}$
[Bra12]	$2^{20I/\epsilon}$	$2^{20I/\epsilon}$
[BBCR10]	$I^{\mathrm{ext}} \times \mathrm{poly}\log(C/\epsilon)$	$I^{\mathrm{ext}} \times \mathrm{poly}\log(C)$

We will prove the first three of these results in this course. But before delving into the proofs let us mention some beautiful consequences of these compression results.

2.1. Information equals amortized communication.

In Table 1 we mentioned that Braverman and Rao [BR14] proved that one can compress a protocol π with k rounds to a protocol τ with worst case communication complexity $O_\epsilon(I + k\log(k) + \sqrt{kI})$ such that the simulation fails with probability at most ϵ. In fact what they prove is stronger. Let us assume for a moment that it is possible to do compression with

$$I + O(k\log(k/\epsilon) + \sqrt{kI}) \tag{1}$$

bits of communication. While this is not true, it is also not far off from the truth. Assuming this, we will give a "pseudo-proof" for the following theorem, and then we will remark how one can turn that into an actual proof using what is actually true.

Theorem 2 (Information equals amortized communication). *Let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$, and let μ be a probability distribution on $\mathcal{X} \times \mathcal{Y}$. For every $\epsilon > 0$, we have*

$$\mathrm{IC}_\mu(f, \mu, \epsilon) = \lim_{n\to\infty} \frac{\mathrm{CC}([f, \mu, \epsilon]^n)}{n}.$$

Semi-proof. One direction is obvious from Theorem 1. Indeed

$$\mathrm{CC}([f, \mu, \epsilon]^n) \geq \mathrm{IC}_\mu([f, \mu, \epsilon]^n) = n\mathrm{IC}_\mu(f, \mu, \epsilon).$$

For the other direction, consider a protocol π that performs $[f, \mu, \epsilon]$. We shall apply compression, and since compression adds error, we have to first decrease the error in π. Indeed by continuity of information there is a protocol π' with information cost $\mathrm{IC}_\mu(\pi') \leq \mathrm{IC}_\mu(\pi) + o_{\delta\to 0}(1)$ that performs $[f, \mu, \epsilon - \delta]$. Let k denote the number of rounds of π, and let τ be the protocol that performs $[f, \mu, \epsilon - \delta]^n$ by running π' on each coordinate separately. Then

$$\mathrm{IC}_{\mu^n}(\tau) = n\mathrm{IC}_\mu(\pi'),$$

while the number of rounds of τ still equals to the number of rounds of π'. Indeed although we are running π on each coordinate separately, we can still bundle all the messages at each round together and send them as one long message.

The rest of this "semi-proof" has a bit of cheating. Using the (not quit right) assumption of (1), we can compress τ to another protocol τ' with

$$\mathrm{CC}(\tau') = n\mathrm{IC}_\mu(\pi') + O(k\log(k/\delta) + \sqrt{kn\mathrm{IC}_\mu(\pi')}),$$

by adding at most δ error. Hence the compressed protocol will perform $[f,\mu,\epsilon]^n$, and thus

$$\lim_{n\to\infty}\frac{\mathrm{CC}([f,\mu,\epsilon]^n)}{n} \le \lim_{\delta\to 0}\lim_{n\to\infty}\frac{n\mathrm{IC}_\mu(\pi') + O(k\log(k/\delta) + \sqrt{kn\mathrm{IC}_\mu(\pi')})}{n} = \lim_{\delta\to 0}\mathrm{IC}_\mu(\pi') = \mathrm{IC}_\mu(\pi).$$

Taking the infimum on the right hand side will prove the result. □

Remark 3. The reason that the above proof is not an actual proof is that one cannot in general compress a protocol to (1) many bits and add only an error of ϵ. However Braverman and Rao [BR14] showed that there is a compression such that if things go right (call it the event E), then the compression perfectly simulates the original protocol, and furthermore will have *expected* communication $I+O(k\log(k/\epsilon)+\sqrt{kI})$, and furthermore E happens (simulation makes no mistake) with probability at least $1-\epsilon$. Now in the above semi-proof, since this compression is applied to many copies, it is possible to apply an argument based on the central limit theorem to show that indeed the *expected* communication will be highly concentrated around its expected value, and thus we can truncate the protocol by terminating it after $I+O(k\log(k/\epsilon)+\sqrt{kI})$ bits of communication to obtain a compressed protocol with the worst case communication $I+O(k\ log(k/\epsilon)+\sqrt{kI})$, and this will fix the above proof. ■

2.2. **The direct sum theorem.** Note that $\mathrm{CC}(f,\mu,\epsilon)=0$ for $f:\mathcal{X}\times\mathcal{Y}\to\mathcal{Z}$ if and only if there exists $z\in\mathcal{Z}$ such that $\mu(f^{-1}(z))\ge 1-\epsilon$.

Theorem 4 (A direct sum theorem). *Let $f:\mathcal{X}\times\mathcal{Y}\to\mathcal{Z}$, and let μ be a probability distribution on $\mathcal{X}\times\mathcal{Y}$. For every $\epsilon>0$, if $\mathrm{CC}(f,\mu,\epsilon)>0$, then for sufficiently large n,*

$$\mathrm{CC}([f,\mu,\epsilon]^n) \ge \frac{\sqrt{n}}{\log^2(n)},$$

and in particular

$$D_\epsilon^{\mu^n}(f^n) \ge \frac{\sqrt{n}}{\log^2(n)}.$$

Proof. Since $\mathrm{CC}(f,\mu,\epsilon)>0$, there exists $\delta>0$ such that $\mathrm{CC}(f,\mu,\epsilon+\delta)>0$ too. Let π be the protocol that performs the task $[f,\mu,\epsilon]^n$ with the smallest communication cost. Obviously

$$\mathrm{CC}(\pi)\le n\mathrm{CC}(f,\mu,\epsilon).$$

Theorem 1 shows that one can use π to obtain a protocol τ for the task $[f,\mu,\epsilon]$ with information cost

$$\mathrm{IC}_\mu(\tau) = \frac{\mathrm{IC}_\mu(\pi)}{n} \le \frac{\mathrm{CC}_\mu(\pi)}{n},$$

and communication cost

$$\mathrm{CC}(\tau) = \mathrm{CC}(\pi) \le n\mathrm{CC}(f,\mu,\epsilon).$$

Now τ has a very small information cost, and we can use the compression of [BBCR10] to compress it to a protocol τ' with communication cost

$$\mathrm{CC}(\tau') = O\left(\sqrt{\mathrm{CC}(\tau) \times \mathrm{IC}_\mu(\tau)} \times \frac{\log(\mathrm{CC}(\tau)/\delta)}{\delta}\right) = \frac{\mathrm{CC}(\pi)}{\sqrt{n}} \times O\left(\frac{\log \mathrm{CC}(\pi)/\delta)}{\delta}\right) = \frac{\mathrm{CC}(\pi)}{\delta\sqrt{n}} \times O(\log n),$$

and error $\epsilon + \delta$. Hence

$$0 < \mathrm{CC}(f, \mu, \epsilon + \delta) \le \frac{\mathrm{CC}(\pi)}{\delta\sqrt{n}} \times O(\log n),$$

which shows

$$\mathrm{CC}(\pi) = \Omega\left(\frac{\sqrt{n}\delta \times \mathrm{CC}(f, \mu, \epsilon + \delta)}{\log n}\right) \ge \frac{\sqrt{n}}{\log^2 n},$$

for sufficiently large n. □

In the above theorem, the term $1/\log^2(n)$ must be ignored as a minor logarithmic fact. Thus we can write

$$\sqrt{n} \lesssim \mathrm{CC}([f, \mu, \epsilon]^n) \le n\mathrm{CC}(f, \mu, \epsilon),$$

where the upper-bound is obvious. How far can one push the lower-bound? It is known that it cannot be pushed all the way up to $\Omega(n\mathrm{CC}(f, \mu, \epsilon))$. Indeed if $\Omega(n)\mathrm{CC}(f, \mu, \epsilon) \le \mathrm{CC}([f, \mu, \epsilon]^n)$, then by dividing both sides by n, and taking the limit, we would get (using information equals amortized communication) that

$$\Omega(\mathrm{CC}(f, \mu, \epsilon)) \le \mathrm{IC}_\mu(f, \mu, \epsilon),$$

and this is not true in general as recently Ganor, Kol, and Raz [GKR14] have shown that $\mathrm{CC}(f, \mu, \epsilon)$ can be exponentially larger than $\mathrm{IC}_\mu(f, \mu, \epsilon)$. However their result does not even overrule the possibility that $(\log \mathrm{CC}(f, \mu, \epsilon))n \le \mathrm{CC}([f, \mu, \epsilon]^n)$.

Problem 5 (Open problem)**.** *Is it possible to improve the $\sqrt{n}$ lower-bound in*

$$\sqrt{n} \lesssim \mathrm{CC}([f, \mu, \epsilon]^n) \le n\mathrm{CC}(f, \mu, \epsilon)?$$

When μ is a product measure, the information cost and the external information cost are equal and hence the compression of [BBCR10] with respect to the external information cost provides a much stronger direct sum theorem.

Theorem 6. *Let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$, and let μ be a* product *probability distribution on $\mathcal{X} \times \mathcal{Y}$. For every $\epsilon > 0$, if $\mathrm{CC}(f, \mu, \epsilon) > 0$, then for sufficiently large n,*

$$\frac{n}{\log^{\Omega(1)}(n)} \le \mathrm{CC}([f, \mu, \epsilon]^n).$$

In particular

$$\frac{n}{\log^{\Omega(1)}(n)} \le D_\epsilon^{\mu^n}(f^n).$$

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

[BR14] Mark Braverman and Anup Rao, *Information equals amortized communication*, IEEE Trans. Inform. Theory **60** (2014), no. 10, 6058–6069. MR 3265014

[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

[FKNN95] Tomás Feder, Eyal Kushilevitz, Moni Naor, and Noam Nisan, *Amortized communication complexity*, SIAM J. Comput. **24** (1995), no. 4, 736–750. MR 1342989 (96j:68089)

[GKR14] Anat Ganor, Gillat Kol, and Ran Raz, *Exponential separation of information and communication*, 55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014, Philadelphia, PA, USA, October 18-21, 2014, 2014, pp. 176–185.

[KRW95] Mauricio Karchmer, Ran Raz, and Avi Wigderson, *Super-logarithmic depth lower bounds via the direct sum in communication complexity*, Comput. Complexity **5** (1995), no. 3-4, 191–204. MR 1394527 (97e:68034)

[Raz98] Ran Raz, *A parallel repetition theorem*, SIAM J. Comput. **27** (1998), no. 3, 763–803 (electronic). MR 1612640 (2000c:68057)

[Sha03] Ronen Shaltiel, *Towards proving strong direct product theorems*, Comput. Complexity **12** (2003), no. 1-2, 1–22. MR 2054892 (2005a:68091)

[Yao82] Andrew C. Yao, *Theory and applications of trapdoor functions*, 23rd annual symposium on foundations of computer science (Chicago, Ill., 1982), IEEE, New York, 1982, pp. 80–91. MR 780384

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