CS 2429 - Foundations of Communication Complexity

Lecture #8: 7 November 2012

Lecturer: Lila Fontes

Today we'll cover some recent results from the paper Lower bounds on information complexity via zero-communication bounds and applications by Kerenidis, Laplante, Lerays, Roland, and Xiao (FOCS 2012).

We'll also recall results covered in past lectures from the papers The partition bound for classical communication complexity and query complexity by Jain and Klauck (CCC 2010) and How to compress interactive communication by Barak, Braverman, Chen, and Rao.

1 Terminology

For P and Q distributions over $\mathbb{X}$, the **statistical distance** of P and Q is:

$$|P - Q| = \max_{S \subseteq \mathbb{X}} \left| \Pr\left[P(S)\right] - \Pr[Q(S)] \right|$$

The **KL divergence** of P and Q is:

$$D(P||Q) = \sum_{x \in \mathbb{X}} P(x) \log \frac{P(x)}{Q(x)}$$

Notice that $D(P||Q) = 0$ when $P \equiv Q$.

For some protocol $\pi : \mathbb{X} \times \mathbb{Y} \to Z$ and distribution μ on $\mathbb{X} \times \mathbb{Y}$, the **information content** is:

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

... where $\pi(X, Y)$ is the transcript of π on X and Y (a random variable) and $I(A; B|C) = \mathrm{H}(A|C) - \mathrm{H}(A|B, C)$ is mutual information (see previous lecture).

Definition A **zero communication protocol** $\pi : \mathbb{X} \times \mathbb{Y} \to \{\perp\} \cup Z$ for some function $f : \mathbb{X} \times \mathbb{Y} \to Z$ works as follows:

- Alice gets $x \in \mathbb{X}$. Bob gets $y \in \mathbb{Y}$.
- Alice performs some computations. Bob performs some computations. They *do not send any messages to each other.* If π has public randomness, they can both see the public random bits.
- Eventually, Alice outputs some $z_a \in Z \cup \{\perp\}$ and Bob outputs some $z_b \in Z \cup \{\perp\}$.
- If $z_a = \perp$ or $z_b = \perp$, then they **abort**, and the protocol's output is $\perp$ (abort). Else, if $z_a = z_b$, the protocol's output is also $\perp$ (abort). Otherwise, the output is z_a.

Of course, our goal in writing zero communication protocols is that when both parties do not abort, Alice and Bob get the same output ($z_a = z_b$). (We want to minimize the chance that neither player aborts, but they disagree on the answer.)

2 Lower bounds

Recall that the partition bound $\text{prt}_\epsilon(f)$ is the optimal value of the following LP.

primal	dual
$\min: \sum_z \sum_R w_{z,R}$	$\max: \sum_{(x,y)\in f^{-1}}(1-\epsilon)\mu_{x,y} + \sum_{(x,y)} \phi_{x,y}$
$\forall (x,y) \in f^{-1}: \sum_{R:(x,y)\in R} w_{f(x,y),R} \geq 1-\epsilon$	$\forall z \forall R: \sum_{(x,y)\in f^{-1}\cap R} \mu_{x,y} + \sum_{(x,y)\in R} \phi_{x,y} \leq 1$
$\forall (x,y): \sum_{R:(x,y)\in R} \sum_z w_{z,R} = 1$	$\forall x \forall y: \mu_{x,y} \geq 0,\ \phi_{x,y} \in \mathbb{R}$
$\forall z \forall R: w_{z,R} \geq 0$	

From previous lectures, we know that lower bounds given by other techniques are dominated by the partition bound:

$$\text{discrepancy bound} \leq \text{rectangle bound} \leq \text{prt}_\epsilon(f) \leq 2^{R^{\text{pub}}_\epsilon(f)}$$

Kerenidis et al. define a new bound, the "relaxed partition bound" $\overline{\text{prt}}_\epsilon(f)$ which sits just below the partition bound in this hierarchy:

$$\text{discrepancy bound} \leq \text{rectangle bound} \leq \overline{\text{prt}}_\epsilon(f) \leq \text{prt}_\epsilon(f) \leq 2^{R^{\text{pub}}_\epsilon(f)}$$

Thus the relaxed partition bound gives stronger lower bounds than all previous techniques (except the partition bound itself).

The relaxed partition bound is defined as the maximum, over all input distributions μ, of the value of the following linear program.

$$\begin{aligned}
&\min_{\eta, p_{R,z}} 1/\eta \text{ subject to:} \\
&\sum_{(x,y)\in f^{-1}} \mu_{x,y} \sum_{R:(x,y)\in R} p_{R,f(x,y)} + \sum_{(x,y)\notin f^{-1}} \mu_{x,y} \sum_{z,R:(x,y)\in R} p_{R,z} \geq (1-\epsilon)\eta \qquad (1) \\
&\forall (x,y) \in \mathbb{X}\times\mathbb{Y}, \sum_{z,R:(x,y)\in R} p_{R,z} \leq \eta \\
&\sum_{R,z} p_{R,z} = 1
\end{aligned}$$

Solutions to the relaxed partition LP can easily be converted to *zero* communication protocols. And zero communication protocols can be converted into solutions of the relaxed partition LP.

The parameter η in the relaxed partition LP corresponds to the *efficiency* of the protocol; η is the probability that the protocol does not abort. Ideally we want this to be as large as possible. Line (1) corresponds to the requirement that the protocol is correct enough (when not aborting).

3 Main theorems

Theorem 1 *There exists a positive constant C such that for all $f: \mathbb{X}\times\mathbb{Y}\to Z$, for all $\epsilon, \delta \in (0, 1/2]$, for all distributions μ on $\mathbb{X}\times\mathbb{Y}$,*

$$\text{IC}_{\mu,\epsilon}(f) \geq \frac{\delta^2}{C}\left(\log \overline{\text{prt}}^{\mu}_{\epsilon+3\delta}(f) - \log|Z|\right) - \delta$$

Thus lower bounds given by the relaxed partition bound — or any weaker technique like discrepancy or the rectangle method – directly translate to lower bounds on information content.[1]

Theorem 2 (Compression) *There exists a universal constant C such that for all distributions μ on inputs, for all communication protocols π, for all $\delta \in (0,1)$, there exists a zero communication protocol π' and real $\lambda \geq 2^{-C(\frac{\mathrm{IC}_{mu}(\pi)}{\delta^2}+\frac{1}{\delta})}$ such that:*[2]

1. $|(\mathbf{X},\mathbf{Y},\pi(\mathbf{X},\mathbf{Y})) - (\mathbf{X},\mathbf{Y},\pi'(\mathbf{X},\mathbf{Y})|\pi'(\mathbf{X},\mathbf{Y}) \neq \perp)| \leq \delta$

   Conditioned on not aborting, the distribution of outputs of π' is close to the distribution of transcripts of π in statistical distance.

3. $\Pr_{r_{\pi'},(x,y)\sim\mu}[\pi'(x,y) \neq \perp] \geq (1-\delta)\lambda$ *The probability that π' does not abort is not* too *small.*

Note in (1) that we are comparing distances of 3-tuples of inputs and *transcripts* of π. The protocol π' will have outputs *transcripts* of π; Alice and Bob will be trying to simulate a likely transcript of π without communicating.

Proof [Theorem 1 using Theorem 2] Let π be a randomized communication protocol achieving $\mathrm{IC}_{\mu,\epsilon}(f)$. Take π' given by Theorem 2. Sample z uniformly from Z using public randomness (this is where the loss of $|Z|$ in efficiency occurs). Sample r_π, the random bits used in protocol π. Let $R = \{x|$ Alice outputs $z\} \times \{y|$ Bob outputs $z\}$. This gives a distribution over labels (R,z) which satisfies $\overline{\mathrm{prt}}^{\mu}_{\epsilon+3\delta}(f)$.

The rest of the lecture will be devoted to Theorem 2.

4 How does zero communication protocol π' work?

In the last lecture, we saw how to compress a communication protocol π to a new protocol π' so that:

$$\mathrm{CC}(\pi') \in O\left(\sqrt{\mathrm{CC}(\pi)\,\mathrm{IC}_\mu\,(\pi)}\frac{\log(\mathrm{CC}(\pi)/\epsilon)}{\epsilon}\right)$$

where the outputs of π' are within statistical distance ϵ of the outputs of π.

This time, we're aiming for $\mathrm{CC}(\pi') = 0$ and statistical distance within some arbitrary small δ. Only the probability of not aborting depends on information content, so somehow this *extreme* compression — down to zero bits of communication! — is going to push a lot of probability into aborting. Protocol π' will abort a lot, depending on the information content $\mathrm{IC}_\mu(\pi)$ of the original protocol π.

4.1 Rejection sampling

In the last lecture, we saw how Alice and Bob could make informed guesses about each other's inputs and use these to shorten the amount of communication required. Now they must make very good guesses indeed.

[1] Prior to this paper, each new lower bound on information for some particular function was *its own publication.* This paper gives a general technique for obtaining lower bounds on information!

[2] Condition (2) is skipped, and has to do with relating π' to the relaxed partiton LP; see the paper for details.

Protocol 3 (Sample solitaire/Zero-communication rejection sampling)

Fix $\pi : \mathbb{X} \times \mathbb{Y} \to \mathbb{U}$ the universe of transcripts. Fix μ a distribution over $\mathbb{X} \times \mathbb{Y}$. Let τ be the distribution of π over $\mathbb{U}$ for inputs drawn from μ. There are public random bits available, so also fix r_π the random bits used in π.

Given $x \in \mathbb{X}$, $y \in \mathbb{Y}$, and parameter Δ, Alice and Bob would like to sample a transcript $u = \pi(x, y)$ according to τ. Neither one knows τ, so they'll have to approximate this as follows:

Setup: For any $u \in \mathbb{U}$,

- *Alice knows $p_A(u)$ the probability of the "Alice bits" in u (where Alice speaks),*
- *Bob knows $p_B(u)$ the probability of the "Bob bits" in u (where Bob speaks),*

and $\tau(u) = p_A(u)p_B(u)$. Alice and Bob don't know τ, but they can build guesses:

- *Alice guesses $\widehat{p}_B(u) = \mathbb{E}_{y\sim\mu}\Pr[\pi(x, y) = u]$ and uses this to build her guess of the overall distribution $\nu_A(u) = p_A(u)\widehat{p}_B(u)$.*
- *Bob guesses $\widehat{p}_A(u) = \mathbb{E}_{x\sim\mu}\Pr[\pi(x, y) = u]$ and uses this to build his guess of the overall distribution $\nu_A(u) = \widehat{p}_A(u)p_B(u)$.*

Draw: They use public coins to sample $u \leftarrow \mathbb{U}$ and $\alpha, \beta \leftarrow [0, 1]$.

Reject: Alice accepts if $\alpha \le p_A(u)/2^\Delta$ and $\beta \le \widehat{p}_B(u)$. Otherwise she rejects ($\perp$).

Bob accepts if $\alpha \le \widehat{p}_A(u)$ and $\beta \le p_B(u)/2^\Delta$. Otherwise he rejects ($\perp$).

Note: The adjustment of 2^Δ essentially means that Alice is allowing for Bob to overestimate p_A by 2^Δ, and Bob is allowing Alice to overestimate p_B by 2^Δ.

What can go wrong? There may be some "bad" transcripts where ν_A or ν_B is far from τ (that is, $\widehat{p}_B$ or $\widehat{p}_A$ is far from p_B or p_A, respectively).

$$B_\Delta(\tau) = \{u \in \mathbb{U} \mid 2^\Delta\nu_A(u) < \tau(u) \text{ or } 2^\Delta\nu_B(u) < \tau(u)\}$$

So for the "good" transcripts $u \notin B_\Delta(\tau)$, both $2^\Delta\widehat{p}_B(u) \ge p_B(u)$ and $2^\Delta\widehat{p}_A(u) \ge p_A(u)$.

Lemma 4 *Let $\gamma = \tau(B_\Delta(\tau))$ be the probability mass of bad transcripts.*

1. $\Pr[\textit{Alice accepts}] = \Pr[\textit{Bob accepts}] = \frac{1}{|\mathbb{U}|2^\Delta}$
2. $\frac{1-\gamma}{|\mathbb{U}|2^{2\Delta}} \le \Pr[\textit{both accept}] \le \frac{1}{|\mathbb{U}|2^{2\Delta}}$
3. *Let τ' be the distribution of transcripts yielded by protocol 3 conditional on both players accepting. Then $|\tau - \tau'| \le \gamma$.*
4. $\forall\tau, \Delta, \epsilon : \tau(B_\Delta(\tau)) \le \frac{D(\tau||\nu_A)+D(\tau||\nu_B)+2}{\Delta}$

Proof

1. $\Pr[\text{Alice} \neq \perp] = \sum_{u\in\mathbb{U}} \frac{1}{|\mathbb{U}|} \cdot \frac{p_A(u)}{2^\Delta} \cdot \widehat{p}_B(u) = \sum_{u\in\mathbb{U}} \frac{\nu_A(u)}{|\mathbb{U}|2^\Delta} = \frac{1}{|\mathbb{U}|2^\Delta}$ and similarly for Bob.

2. If $u \notin B_\Delta(\tau)$ then $2^\Delta \widehat{p}_B(u) \geq p_B(u)$ and $2^\Delta \widehat{p}_A(u) \geq p_A(u)$. So if α passes Alice's check, it passes Bob's check too, and similarly for β.

$$\Pr[\text{both accept } u] = \frac{1}{|\mathbb{U}|} \cdot \frac{p_A(u)}{2^\Delta} \cdot \frac{p_B(u)}{2^\Delta} = \frac{\tau(u)}{|\mathbb{U}|2^{2\Delta}} \tag{2}$$

This gives an easy lower bound on the probability they both accept.

$$\Pr[\text{both accept}] = \sum_{u \in U} \Pr[\text{both accept } u] \geq \sum_{u \notin B_\Delta(\tau)} \Pr[\text{both accept } u] = \frac{1-\gamma}{|\mathbb{U}|2^{2\Delta}}$$

We can also find an upper bound better than just squaring the probability from the first part of the lemma, since Alice and Bob are not acting entirely independently.

$$\begin{aligned}
\Pr[\text{both accept}] &= \sum_{u \in \mathbb{U}} \Pr[\text{both accept } u] \\
&= \sum_{u \in \mathbb{U}} \frac{1}{|\mathbb{U}|2^{2\Delta}} \cdot \min(p_A(u)/2^\Delta, \widehat{p}_A(u)) \cdot \min(p_B(u)/2^\Delta, \widehat{p}_B(u)) \\
&\leq \sum_{u \in \mathbb{U}} \frac{\tau(u)}{|\mathbb{U}|2^{2\Delta}} = \frac{1}{|\mathbb{U}|2^{2\Delta}}
\end{aligned}$$

3. Let $\Pr[\text{both accept}] = \eta \in [\frac{1-\gamma}{|\mathbb{U}|2^{2\Delta}}, \frac{1}{|\mathbb{U}|2^{2\Delta}}]$. We want that $\forall S \subseteq \mathbb{U}$, $\tau(S) - \tau'(S) \leq \gamma$.

$$\begin{aligned}
\tau(S) - \tau'(S) &= \tau(S \cap B_\Delta(\tau)) - \tau'(S \cap B_\Delta(\tau)) + \tau(S \cap \overline{B_\Delta(\tau)}) - \tau'(S \cap \overline{B_\Delta(\tau)}) \\
&\leq \gamma - 0 + \tau(S \cap \overline{B_\Delta(\tau)}) - \tau'(S \cap \overline{B_\Delta(\tau)}) \\
&= \gamma + \tau(S \cap \overline{B_\Delta(\tau)}) - \frac{\tau(S \cap \overline{B_\Delta(\tau)}}{|\mathbb{U}|2^{2\Delta}\eta} && \text{by equation (2) above} \\
&= \gamma + \tau(S \cap \overline{B_\Delta(\tau)})\big(1 - \frac{1}{|\mathbb{U}|2^{2\Delta}\eta}\big) \\
&\leq \gamma && \text{because } \eta \leq \frac{1}{|\mathbb{U}|2^{2\Delta}}
\end{aligned}$$

4. Homework. Alternatively, available in Mark Braverman's Interactive information complexity (STOC 2012).

4.2 Assembling π'

Is protocol 3 good enough to get our zero-communication protocol? It has good correctness (condition 3) but its efficiency is too small (condition 2).

In particular, we'd like to eliminate the dependence on $|\mathbb{U}|$.

There is some correlation between Alice not aborting and Bob not aborting, so we will have them repeat protocol 3 many times. Then Alice and Bob (still without communicating!) will try ty agree on which of the trials to pick.

This ends up being enough to eliminate the dependence on $|\mathbb{U}|$, although it will add some complication.

(In notes to follow, many extraneous constants have been removed to simplify the reading experience. Refer to original paper for precise statements.)

Protocol 5 (Zero communication protocol π')

Everyone knows the protocol π, μ the distribution on inputs, $\delta > 0$, and $\mathrm{IC}_\mu(\pi) > 0$.
Alice gets $x \in \mathbb{X}$, Bob gets $y \in \mathbb{Y}$. The protocol π' proceeds as follows:

Setup: *Alice builds p_A, $\widehat{p}_B$ as before. Bob builds p_B, $\widehat{p}_A$.*

Set $\Delta = I/\delta^2$, $T = |\mathbb{U}|2^\Delta \ln(1/\delta)$, and $k = \Delta + \log\left(\frac{\log^2(\delta)}{\delta}\right)$.

Sample: *Repeatedly sample transcripts (using protocol 3) T times.*

Alice produces a list $(a_1, a_2, \ldots, a_T) \in (\mathbb{U} \times \{\perp\})^T$, Bob produces a list $(b_1, b_2, \ldots, b_T)$.

(These lists contain many $\perp$s, but if $a_i \neq \perp$ then $a_i = b_i$.)

Coordinate: *Use public coins to fix $h : [T] \to \{0,1\}^k$ and $r \in \{0,1\}^k$.*

Alice finds the smallest i such that $a_i \neq \perp$. If $h(i) = r$, Alice outputs a_i; otherwise she aborts.

Bob finds the smallest j such that $b_j \neq \perp$ and $h(j) = r$. If there is none, he aborts; otherwise, he outputs that b_j.

Lemma 6 *Let D be the event that the inputs have large divergence: $D(\tau||\eta_A)$ or $D(\tau||\eta_B) > I/\delta$. Let C be the event that there is a collision: $\exists i \neq j$ such that $h(i) = h(j) = r$.*

1. *Inputs rarely have large divergence:* $\Pr_{(x,y)\sim\mu}[D] \leq \delta/4$
2. *Collisions are rare:* $\forall x \forall y \Pr_{r_{\pi'}}[C] \leq \delta/2^{k+\Delta}$
3. *π' does not abort too much (when the divergence is good):*
   $\forall x \forall y$ *if* $\neg D$ *then* $\Pr_{r_{\pi'}}[\pi'(x,y) \neq \perp] \geq \frac{1-\delta}{2^{k+\Delta}}$.
5. *$\forall \pi \forall \mu \forall \delta$, π' satisfies: $\forall x \forall y$ such that $\neg D$, let $\pi'_{\not\perp}(\mathbf{X}, \mathbf{Y})$ be the distribution of $\pi'(\mathbf{X}, \mathbf{Y})$ conditioned on not aborting. Then:*

$$|\pi(\mathbf{X}, \mathbf{Y}) - \pi'_{\not\perp}(\mathbf{X}, \mathbf{Y})| \leq \mathbf{3}\delta/\mathbf{4}$$

For proof of Lemma 6 (and its omitted parts), see the paper.

The zero communication protocol π' given by protocol 5 is sufficient to prove the compression theorem.

Proof [Theorem 2 using Lemma 6]

1. Part 1 follows from parts 1 and 5 of the lemma.
3. Part 3 follows from parts 1 and 3 of the lemma.