LECTURE 10

Today (and next lecture):
Linear lower bound for $DISJ_n$ in 2-party randomized CC.

Proof from [BJKS04] using information theory.

Let us start by high-level (and incorrect) outline of what a proof might look like. This won't work, but it will give intuition and we will fix the broken details later.

Recall in lecture 8 we talked briefly about direct-sum. Suppose we have function g. How much does it cost to compute n independent copies of g? Does the cost scale linearly with n?

For $x = (x_1, \ldots, x_n) \in \{0,1\}^n$, $y = (y_1, \ldots, y_n) \in \{0,1\}^n$
$DISJ_n(x,y) = \neg \bigvee_{i=1}^{n} (x_i \wedge y_i)$

[Some confusion as to whether value should be 0 or 1 for $x \cap y = \emptyset$. BJKS has 0 but we will try to stick to 1 as before]

So intuitively, any protocol for $DISJ_n$ solves n copies of (bitwise) AND.
Suppose that we could show:

(a) cost of n copies $\geq n \cdot$ cost of 1 copy
(b) cost of 1 copy is $\Omega(1)$

Then get LB $\Omega(n)$.

Don't know how to do this...

Look instead at mutual information of inputs (X, Y) and protocol $\Pi(X, Y)$. A protocol can't convey more info than the # bits communicated, so information lower bounds yield CC lower bounds.

Suppose X & Y sampled independently with 0/1 with equal probability in each coordinate.
Write distribution $(X, Y) = (X_1, Y_1, X_2, Y_2, \ldots, X_n, Y_n)$

Since all coordinates are independent, the chain rule for information mutual yields

$$(*) \quad I((X,Y); \Pi(X,Y)) = \sum_{i=1}^{n} I((X_i, Y_i); \Pi(X,Y))$$

This is a direct-sum-kind-of result!
So if we can prove that protocol Π has to reveal same info about each pair (x_i, y_i) of coordinates, then get linear lower bound!

This won't work either...

Because for inputs drawn from this distribution, $DISJ_n(x,Y)$ ~~is~~ is exceedingly likely to be zero (Why?).

Need to look at mutual information w.r.t more complicated distributions.
But then independence need not hold, and if so $(*)$ fails...

But we will find a way to fix all this! (Or rather, BJKS will...)

Quick detour: public coins vs. private coins

So far studied mostly public-coin protocols
Today will do private-coin protocols.

The latter are weaker!

$$R^{priv}_{\varepsilon}(EQ_n) = \Theta(\log n)$$ [follows from what we said in class]

$$R^{pub}_{\varepsilon}(EQ_n) = O(1)$$ [did in class]

But the gap can't be larger than that

THM For any $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ any any $\varepsilon, \delta > 0$ it holds that

$$R^{priv}_{\varepsilon+\delta}(f) \leq R^{pub}_{\varepsilon}(f) + O\left(\log n + \log(1/\delta)\right)$$

Won't do the proof since we want to use time for other stuff
But this means that if we want to prove linear lower bounds for $DISJ_n$, private or public coins really doesn't matter.

~~~

Notation (ν Greek letter nu)

X for random variable

$X \sim \nu$ X distributed according to ν

$\vec{X}$ vector random variable (skip arrow when clear from context)

$\vec{x}, \vec{y}$ range over domains $\mathcal{X}^n, \mathcal{Y}^n$

Suppose $\vec{X} = (X_1, \ldots, X_m) \sim \mu$. Then μ is a PRODUCT DISTRIBUTION if all X_i are mutually independent.

Ex Let $\vec{X} = (X_1, \ldots, X_m)$ be the result of m independent coin flips
~~~

Recap of information theory

Let μ (Greek letter mu) distribution on finite set Ω

Let $X \sim \mu$. $\mu(\omega) = \Pr[X = \omega]$

Entropy $H(X) = \sum_{\omega \in \Omega} \mu(\omega) \log\left(\frac{1}{\mu(\omega)}\right)$

Conditional entropy

$$H(X|Y) = \sum_{y} \Pr[Y=y] \cdot H(X \mid Y=y)$$

$H(X \mid Y=y)$ entropy of conditional distribution of X given event $Y=y$

Joint entropy of X, Y is entropy of joint distribution (X,Y), denoted $H(X,Y)$

Mutual information between X and Y

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

Conditional mutual information between X and Y conditional on Z

$$I(X;Y|Z) = H(X|Z) - H(X \mid Y,Z)$$

$$= \sum_{z} \Pr[Z=z] \cdot I(X;Y \mid Z=z)$$

"How much info does Y reveal about X"

$H(X)$ = uncertainty of X

$H(X|Y)$ = uncertainty of X given Y

$I(X,Y) = H(X) - H(X|Y)$ = how much did uncertainty decrease
= how much info did Y reveal

Proposition 1 (recap)

1) $0 \leq \mathcal{H}(X) \leq \log |\mathcal{X}|$

2) $I(X;Y) \geq 0$

3) Subadditivity of entropy

$\mathcal{H}(X,Y) \leq \mathcal{H}(X) + \mathcal{H}(Y)$ with equality iff X and Y are independent.

4) Subadditivity of conditional entropy

$\mathcal{H}(XY|Z) \leq \mathcal{H}(X|Z) + \mathcal{H}(Y|Z)$ with equality iff X and Y are independent conditioned on Z.

5) Data processing inequality

If X, Z conditionally independent given Y, then $I(X;Y|Z) \leq I(X;Y)$.

Suppose $\mathcal{K} \subseteq \mathcal{X}^n \times \mathcal{Y}^n$ set of inputs to f

$f : \mathcal{K} \to \{0,1\}$ function.

For $DISJ_n$, $\mathcal{X} = \mathcal{Y} = \{0,1\}$

$x, y \subseteq [n]$ or equivalently $\in \{0,1\}^n$

$DISJ_n(x,y) = 1$ iff $x \cap y = \emptyset$

Def 2 Information cost of protocol

Π randomized protocol inputs from $\mathcal{K}$.

μ distribution on $\mathcal{K}$, $(X,Y) \sim \mu$.

The INFORMATION COST of Π w.r.t. μ is

$$I((X,Y); \Pi(X,Y)).$$

Ex 3 Intuitively, the most info any protocol can give is to output x & y. So information cost $\leq H(X,Y)$.

$I(X,Y;\ \Pi(X,Y)) \geq 0$ since info is non negative

$= 0$ iff protocol transcript independent of inputs $\Rightarrow$ if so, will be very hard to compute any non-constant function f.

So we would expect that a protocol has to reveal some info about inputs.

Def 4 Information complexity of function

The δ-error INFORMATION COMPLEXITY of f w.r.t. distribution μ, denoted $IC_{\mu,\delta}(f)$, is the minimum information cost of any δ-error protocol for f wrt μ.

Proposition 5

For any distr μ and any error δ

$$R_\delta^{pub}(f) \geq IC_{\mu,\delta}(f)$$

Proof Let Π denote best protocol for f in terms of communication. Let $|\Pi|$ length of longest transcript. Suppose $(X,Y) \sim \mu$.

Then
$$\begin{aligned} R_\delta^{pub}(f) &= |\Pi| \\ &\geq H(\Pi(X,Y)) \\ &\geq I(X,Y;\ \Pi(X,Y)) \\ &\geq IC_{\mu,\delta}(f) \end{aligned}$$

Suppose $f : \mathcal{X}^n \times \mathcal{Y}^n \to \{0,1\}^n$ can be described in terms of simpler function $h : \mathcal{X} \times \mathcal{Y} \to \{0,1\}^k$, so that

$$f(x,y) = f(h(x_1,y_1), h(x_2,y_2), \ldots, h(x_n,y_n))$$

Ex $DISJ_n(x,y) = \neg \bigvee_{i=1}^{n} (x_i \wedge y_i)$

If f depend symmetrically on each coordinate, then would expect that any protocol for f must (implicitly) solve all subinstances $h(x_i, y_i)$. Furthermore, if μ product distribution, can hope to use chain rule to prove

$$IC(f) \geq n \cdot IC(h)$$

But μ might not be product distribution; in particular X and Y might not be independent. Leads to next def

<u>Def 6</u> Mixture of product distributions

Let ν distribution on $\mathcal{X} \times \mathcal{Y}$.
Let T random variable over $\mathcal{T}$ such that conditioned on T, X and Y are independent.
Let η be joint distribution of $((X,Y),T)$.
Then η is a MIXTURE OF PRODUCT DISTRIBUTIONS

(η: Greek letter eta)

Looking at just (X,Y), the marginal distribution is still ν.

Observation 7

Suppose ν on $\mathcal{X} \times \mathcal{Y}$ non-product, $\mu = \nu^n$ (i.e., n independent copies of ν).
Let $(X, Y) \sim \nu$.
Suppose T is random variable s.t. X and Y are independent conditioned on T.
Let η joint distribution on $((X,Y), T)$
Let $\zeta = \eta^n$. (Greek letter zeta)

Then ζ is a mixture of product distributions with marginal distribution $\mu = \nu^n$ on $\vec{\mathcal{X}} \times \vec{\mathcal{Y}}$.
Furthermore, for $((\vec{X}, \vec{Y}), \vec{T}) \sim \zeta$ it holds that all (X_j, Y_j), $j \in [n]$ are mutually independent of each other, and this continues to hold even when conditioned on $\vec{T}$

Ex 8 Consider ν on $\{0,1\}^2$ defined by

(x,y)	$\nu(x,y)$
$(0,0)$	$1/2$
$(0,1)$	$1/4$
$(1,0)$	$1/4$
$(1,1)$	0

ν not a product distr.

Now let T be uniform on $\{A, B\}$

If $T = A$ then let $X = 0$, Y uniform on $0,1$
If $T = B$ -"- X uniform on $\{0,1\}$, $Y = 0$

Then X, Y independent conditioned on T
Also $(X,Y) \sim \nu$. Hence $((X,Y), T)$ mixture

Def 9 CONDITIONAL INFORMATION COST

Let Π randomized protocol with inputs from $\mathcal{K} \subseteq \mathcal{X}^n \times \mathcal{Y}^n$. Suppose $((\vec{X},\vec{Y}),\vec{T}) \sim \zeta$ and that ζ is a mixture of product distributions on $\mathcal{K} \times \mathcal{T}$. Then the CONDITIONAL INFORMATION cost of Π w r t ζ is

$$I(\vec{X},\vec{Y}; \Pi(\vec{X},\vec{Y}) \mid \vec{T}).$$

Def 10 CONDITIONAL INFORMATION COMPLEXITY

The δ-error CONDITIONAL INFORMATION COMPLEXITY of f w.r.t ζ, denoted $CIC_{\zeta,\delta}(f)$, is the minimum conditional information cost of a δ-error protocol for f w r t ζ.

Prop 11

Let $f: \mathcal{K} \to \{0,1\}$ function
μ distribution on $\mathcal{K}$
ζ mixture of product distributions on $\mathcal{K} \times \mathcal{T}$ such that marginal distribution on $\mathcal{K}$ is μ.

Then $$IC_{\mu,\delta}(f) \geq CIC_{\zeta,\delta}(f)$$

Proof Let Π protocol with info cost $IC_{\mu,\delta}(f)$.
Let $((\vec{X},\vec{Y}),\vec{T}) \sim \zeta$, where $(X,Y) \sim \mu$.

$\Pi(\vec{X},\vec{Y})$ is conditionally independent of $\vec{T}$ given $\vec{X},\vec{Y}$ (coin flips of Π are independent of T.

By the data processing inequality

$$IC_{\mu,\delta}(f) = I(\vec{X},\vec{Y}; \Pi(\vec{X},\vec{Y})) \geq I(\vec{X},\vec{Y}; \Pi(\vec{X},\vec{Y}) \mid \vec{T}) \geq CIC_{\zeta,\delta}(f)$$

Corollary 12

Suppose $f : \mathcal{X} \to \{0,1\}$

ζ mixture of product distributions on $\mathcal{X} \times \mathcal{Y}$

Then $R_\delta^{pub}(f) \geq CIC_{\zeta,\delta}(f)$

DIRECT SUM FOR CONDITIONAL INFORMATION CPLX

Now time to

- formally define what we mean by decomposing function into primitive functions
- prove direct sum theorem

[We follow notation from BJKS in case you want to read the paper and compare.]

Now think of domain $\mathcal{X} = \mathcal{L}^n$ $\mathcal{L} \subseteq \mathcal{X} \times \mathcal{Y}$

$f : \mathcal{L}^n \to \{0,1\}$ Think $\mathcal{L} \subseteq \{0,1\} \times \{0,1\}$

Π δ-error protocol for f

ζ mixture of product distributions on $\mathcal{L} \times \mathcal{O}$

Let $\eta = \zeta^n$

Suppose $((\vec{X}, \vec{Y}), \vec{O}) \sim \eta$

Unless stated otherwise, always assume the above in lemmas & theorems below

~~~

NB! switching roles of η and ζ

Changing from T to D

This is to be consistent with (inconsistent) BJKS-notation
~~~

Lemma 13 Information cost decomposition lemma

(With assumptions as above)

$$I(\vec{X},\vec{Y};\pi(\vec{x},\vec{y})\mid\vec{\mathcal{D}}) \geq \sum_{j=1}^{n} I(X_j;Y_j;\pi(\vec{x},\vec{y})\mid\vec{\mathcal{D}})$$

Proof Write $\pi(\vec{x},\vec{y}) = \pi$ for brevity.

By definition $I(\vec{X},\vec{Y};\pi\mid\vec{\mathcal{D}}) = H(\vec{X},\vec{Y}\mid\vec{\mathcal{D}}) - H(\vec{X},\vec{Y}\mid\pi,\vec{\mathcal{D}})$ (*)

X_j, Y_j independent conditioning on $\vec{\mathcal{D}}$. Hence

$$H(\vec{X},\vec{Y}\mid\vec{\mathcal{D}}) = \sum_j H(X_j,Y_j\mid\vec{\mathcal{D}}) \qquad (1)$$

By subadditivity of conditional entropy

$$H(\vec{X},\vec{Y}\mid\pi,\vec{\mathcal{D}}) \leq \sum_j H(X_j,Y_j\mid\pi,\vec{\mathcal{D}}) \qquad (2)$$

Plug (1) & (2) into definition (*), recede terms, and use def of conditional mutual info again. ☐

Def 14 $f: \mathcal{X}^n \to \{0,1\}$ is g-DECOMPOSABLE WITH PRIMITIVE h if

$$f(\vec{x},\vec{y}) = g(h(x_1,y_1),\ldots,h(x_n,y_n))$$

Ex 15 $DISJ_n(\vec{x},\vec{y}) = \neg \bigvee_{i=1}^{n} (x_i \wedge y_i)$

Ex 16 Inner product of $\vec{x}$ and $\vec{y}$ on $\{0,1\}^n = \mathbb{F}_2^n$ is $\langle\vec{x},\vec{y}\rangle = \sum_i x_i \cdot y_i \pmod 2$

Let $h(x_i,y_i) = x_i \wedge y_i$

$g(z_1,\ldots,z_n) = \bigoplus_{i=1}^{n} z_i$

Now we want to lower bound info about each coordinate pair in $\vec{x}, \vec{y}$ by conditional info cplx of h

In symbols $I(x_j, y_j ; \Pi \mid \mathcal{D}) \geq CIC_{\zeta,\delta}(h)$.

Do this by "forcing Π to solve h in each coordinate"

Pick distribution so that values of x_j and y_j determine everything

<u>Def 17</u> Embedding

$\vec{w} \in \mathcal{L}^n$, $j \in [n]$, $u \in \mathcal{L}$

$$\text{EMBED}(\vec{w}, j, u) = \begin{cases} w_i & \text{for } i \neq j \\ u & \text{for } i = j \end{cases}$$

i.e., replace jth component in $\vec{w}$ by u; leave the rest intact.

<u>Def 18</u> Collapsing distribution

Suppose $f : \mathcal{L}^n \to \{0,1\}$ g-decomposable with promise $h : \mathcal{L} \to \{0,1\}$.

$(\vec{x}, \vec{y}) \in \mathcal{L}^n$ is a COLLAPSING INPUT for f if for every $j \in [n]$, every $(u,v) \in \mathcal{L}$

$$f(\text{EMBED}(\vec{x}, j, u), \text{EMBED}(\vec{y}, j, v)) = h(u,v)$$

μ on $\mathcal{L}^n$ is a COLLAPSING DISTRIBUTION for f if every $(\vec{x}, \vec{y})$ in support of μ is a collapsing input.

Actually, with our def of DIST we will have $\neg h(u,v) = 1 - h(u,v)$, but this clearly doesn't matter

Ex 19 Distribution $\mu = \nu^n$ for ν as in Ex 8 is collapsing, since $\nu(1,1) = 0$

In this way, protocol Π for f yields n protocols for h: Compute $h(u,v)$ by embedding (u,v) in j'th coordinates for $j \in [n]$.

(At most) same error as Π.

Lemma 20 Reduction Lemma

Let ζ mixture of product distributions on $\mathcal{L} \times \mathcal{D}$.

$\eta = \zeta^n$ $\quad ((\vec{X}, \vec{Y}), \vec{D}) \sim \eta$

If distribution on $\vec{X}, \vec{Y}$ is collapsing for f, then

$$I(X_j, Y_j ; \Pi(\vec{X}, \vec{Y}) \mid \vec{D}) \geq CIC_{\zeta,\delta}(h)$$

for all $j \in [n]$

This requires a proof, which we skip for now.

Theorem 21 Direct sum theorem

Let $f : \mathcal{L}^n \to \{0,1\}$ decomposable with primitive h.

Let ζ mixture of product distributions on $\mathcal{L} \times \mathcal{D}$

Let $\eta = \zeta^n$, $((\vec{X}, \vec{Y}), \vec{D}) \sim \eta$

If distribution of $(\vec{X}, \vec{Y})$ is collapsing for f, then

$$CIC_{\eta,\delta}(f) \geq n \cdot CIC_{\zeta,\delta}(h).$$

Proof [of Thm 21]

Let Π optimal δ-error protocol in terms of conditional info cost wrt η.

Then $CIC_{\eta,\delta}(f) = I(\vec{X},\vec{Y}; \Pi(\vec{X},\vec{Y}) \mid \vec{D})$.

By info cost decomposition Lemma (Lem 13)

$$I(\vec{X},\vec{Y}; \Pi(\vec{X},\vec{Y}) \mid \vec{D}) \geq \sum_j I(X_j, Y_j; \Pi(\vec{X},\vec{Y}) \mid \vec{D}) \quad (\dagger)$$

By reduction Lemma (Lem 20) we have

$$(\dagger) \geq n \cdot CIC_{\zeta,\delta}(h). \qquad \square$$

Corollary 22

$$R_\delta^{priv}(f) \geq CIC_{\eta,\delta}(f) \geq n \cdot CIC_{\zeta,\delta}(h).$$

Thus, for set disjointness we have

$$R_\delta^{priv}(DISJ_n) \geq n \cdot CIC_\zeta(AND)$$

Hence sufficient to prove $\Omega(1)$ lower bound for conditional information complexity of 1-bit AND wrt ζ.