1 A short recap of last lecture

1.1 Some properties of information quantities

\textbf{Theorem 1} (Chain rule for Entropy).

\[ H(X,Y) = H(X) + H(Y|X) \]

\textbf{Theorem 2} (Conditioning reduces entropy).

\[ 0 \leq H(Y|X) \leq H(Y). \]
Moreover, \( H(Y|X) = H(Y) \) iff \( X,Y \) are independent, and \( H(Y|X) = 0 \) iff \( Y \) is determined by \( X \), in other words \( Y = f(X) \) for some function \( f \).

We have defined the mutual information of two random variables:

\[
I(X,Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)
\]

**Theorem 3** (Chain rule for Mutual Information).

\[
I(X,Y;Z) = I(X;Z) + I(Y;Z|X).
\]

As a warm up we will prove the following claim:

**Claim 1.1.** Let \( A,B,C \) be random variables such that \( A,C \) are independent. Then:

\[
I(A;B) \leq I(A;B|C)
\]

**Proof.**

\[
I(A;B) = H(A) - H(A|B)
\leq H(A) - H(A|BC) \quad \text{(conditioning reduces entropy)}
= H(A|C) - H(A|BC) \quad \text{(A, C are independent)}
= I(A;B|C).
\]

One extreme example for such \( A,B,C \) is when \((A,B,C)\) is a uniformly random vector in \(\{0,1\}^3\) with an even number of 1’s. The details of this example are left for the reader.

### 1.2 External and Internal Information of a protocol

Let \( \Pi \) be a protocol and let \( \mu \) be a distribution on the input space \( \mathcal{X} \times \mathcal{Y} \). We defined **The External Information of \( \Pi \) w.r.t \( \mu \)**:

\[
\text{IC}_{\mu}^{\text{ext}}(\Pi) = I(\Pi;X,Y).
\]

We argued that this quantity measures how much do an external observer who knows \( \mu \), learns on the input from the transcript of the protocol. We proved that for all \( \mu \) the external information of a protocol bounds from below its average communication complexity (w.r.t \( \mu \)).

Finally, we defined **The Internal Information of \( \Pi \) w.r.t \( \mu \)**:

\[
\text{IC}_{\mu}^{\text{int}}(\Pi) = I(\Pi;X|Y) + I(\Pi;Y|X).
\]
We argued that it measures “How many bits does Alice learns\(^1\) on Bob input by reading the transcript given what she knows plus how many bits does Bob learns on Alice input by reading the transcript given what he knows”.

### 1.2.1 A discussion on the definition of Internal Information

Another natural way to define Internal Information of a protocol \(\Pi\) is \(I(\Pi; X|Y, R_b) + I(\Pi; Y|X, R_a)\). Moreover, on an intuitive level, knowing \(R_a\) \((R_b)\) should not change the amount of information \(\Pi\) teaches Alice (Bob) on \(Y\) \((X)\) and therefore one would expect that:

\[
I(\Pi; X|Y, R_b) + I(\Pi; Y|X, R_a) = I(\Pi; X|Y) + I(\Pi; Y|X).
\]

In a similar manner to how we proved Claim 1.1 it is possible to show that \(I(\Pi; X|Y, R_b) + I(\Pi; Y|X, R_a) \geq I(\Pi; X|Y) + I(\Pi; Y|X)\). One way to make these definitions equivalent is to make the following technical modification in the definition of private coin protocols: At each round each of the parties uses a private random string which is independent of the private strings of the other rounds. With this modification, standard induction yields the desired equivalence. Why do we call this modification “technical”? The reason is that it doesn’t change the “power” of protocols in the following sense: For every protocol \(\Pi\), there exists a modified protocol \(\Pi'\) such that for every input \((x, y)\), the distributions \(\Pi(x, y)\) and \(\Pi'(x, y)\) are the same. This means that whatever function with whatever information/communication complexities one can compute with “standard” protocols, one can also compute with modified protocols. Infact, I think that all the examples of private coin protocols that we have seen in class were “modified” protocols.

### 1.2.2 Internal Information \(\leq\) External Information

Internal Information measures how much do the parties learn on each others input from the transcript of the protocol and their own input. External Information measures how much do an external observer who has no prior information about the input learns about the input from the transcript of the protocol. Thus, it is intuitive to expect that the internal information is lesser or equal from the external information. The next Lemma proves this fact. The idea of the proof is to use induction and to show that each message in the transcript teaches an external observer more than it teaches the parties about the input.

**Lemma 1.2.** For every protocol \(\Pi\) and for every input distribution \(\mu\):

\[
IC^\text{int}_\mu (\Pi) \leq IC^\text{ext}_\mu (\Pi)
\]

**Proof.** We prove this statement by induction on the number of rounds in \(\Pi\). The case of 0 rounds is trivial.

\(^1\)also in this case, Alice and Bob know \(\mu\)
Let \( \Pi \) be a protocol with at least 1 round. Let \( M_1 \) be the (random) message sent in the first round. Assume without loss of generality that \( M_1 \) is sent by Alice.

\[
\text{IC}^\text{int}_\mu (\Pi) = I(\Pi; X|Y) + I(\Pi; Y|X)
\]

\[= I(\Pi; M_1; X|Y) + I(\Pi; M_1; Y|X) \quad (M_1 \text{ is determined by } \Pi)\]

\[= [I(M_1; X|Y) + I(M_1; Y|X)] + [I(\Pi; X|Y, M_1) + I(\Pi; Y|X, M_1)] \quad (by \text{the Chain rule})\]

\[\leq [I(M_1; X|Y) + I(M_1; Y|X)] + [I(\Pi; X, Y|M_1)]\]

(By the induction hypothesis applied on the subprotocols of \( \Pi \) after the first message)

\[= [I(M_1; X|Y)] + [I(\Pi; X, Y|M_1)] \quad (M_1 \text{ is sent by Alice and thus it is independent of } Y \text{ given } X)\]

\[\leq [I(M_1; X, Y)] + [I(\Pi; X, Y|M_1)] \quad (by \text{the chain rule: } I(M_1; X|Y) \leq I(M_1; X, Y))\]

\[= \text{IC}^\text{ext}_\mu (\Pi) \quad (By \text{the chain rule})\]

\[
\text{Exercise 1.3. } \text{Prove that if } \mu \text{ is a product distribution then for every protocol } \Pi:\]

\[
\text{IC}^\text{int}_\mu (\Pi) = \text{IC}^\text{ext}_\mu (\Pi).
\]

(Hint: Show that for any three random variables \( X, Y, Z \), if \( X, Y \) are independent then \( I(X, Y; Z) \leq I(X; Z|Y) + I(Y; Z|X) \).)

### 2 Direct Sum for Internal Information

#### 2.1 Distributional Internal Information of a function

In the rest of the lecture we will deal with the following complexity measure: Given a function \( f : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{Z} \) and \( \mu \), a distribution on the domain of \( f \), we define the distributional Internal Information of \( f \) w.r.t \( \mu \):

\[
\text{IC}^\text{int}_{\mu, \epsilon} (f) = \min_{\Pi} \left( \text{IC}^\text{int}_\mu (\Pi) \right).
\]

Where \( \Pi \) is quantified over all randomized protocols which compute \( f \) with \( \epsilon \) error; note that this means that \( \Pi \) has to compute \( f \) with \( \epsilon \) error regardless of \( \mu \); namely, for every \( x, y \), the probability (over the randomness of \( \Pi \)) that \( \Pi \) errs is at most \( \epsilon \). It might take some time to get used to this setup since when we analyzed communication complexity we usually had either deterministic protocols with a distribution on the input or randomized protocols with no distribution on the input and Yao Min-Max principle connected between the two. Here we have a mix: Distribution on the input and randomized protocols.

As an example consider the AND function and the distribution:
In the definition of distributional Internal Information, if we would have quantified over all protocols that compute \( AND \) with \( \epsilon \) error on this particular distribution, then the protocol which always outputs 0 would be a “kosher” protocol that have 0 internal information. However, we quantify over protocols which always compute \( f \) with \( \epsilon \)-error and therefore this protocol is not “kosher”. We will later see that for the above distribution, the distributional internal information of \( AND \) is \( > 0 \).

2.2 Direct Sum

Let \( f : \mathcal{D}' \to \mathcal{R}' \) and \( g : \mathcal{D}'' \to \mathcal{R}'' \) be two functions. We define \( f \times g : \mathcal{D}' \times \mathcal{D}'' \to \mathcal{R}' \times \mathcal{R}'' \) by

\[
(f \times g)(d', d'') = (f(d'), g(d'')), \quad \forall d' \in \mathcal{D}', d'' \in \mathcal{D}''.
\]

We say that a protocol \( \Pi \) is a marginal \( \epsilon \)-error protocol for \( f_1 \times f_2 \) if for every \( i \in \{1, 2\} \) and for every \((x_1, y_1), (x_2, y_2) \) \( \in \) Dom \( (f_1) \times \) Dom \( (f_2) \), the probability that \( \Pi \) \((x_i, y_i)\) errs on \( (x_i, y_i) \) is at most \( \epsilon \). Note that every \( \epsilon \)-error protocol for \( f \times g \) is also a marginal \( \epsilon \)-error protocol for \( f \times g \) and that every marginal \( \epsilon \)-error for \( f \times g \) is a \( 2 \epsilon \)-error protocol for \( f \times g \).

**Lemma 2.1** (The Decomposition Lemma for Internal Information). Let \( \Pi_{f \times g} \) be a marginal \( \epsilon \)-error protocol for \( f \times g \), and let \( \mu_f \times \mu_g \) be a product distribution on \( \text{Dom} \( (f) \times \text{Dom} \( (g) \) \). Then there exist \( \epsilon \)-error protocols \( \Pi_f, \Pi_g \) for \( f, g \) respectively such that:

\[
\text{IC}^\text{int}_{\mu \times \mu}(\Pi_{f \times g}) \geq \text{IC}^\text{int}_{\mu_f}(\Pi_f) + \text{IC}^\text{int}_{\mu_g}(\Pi_g)
\]

**Proof.** Let \( \Pi \) denote the random transcript of \( \Pi_{f \times g} \) when the input is distributed like \( \mu_f \times \mu_g \)

\[
\text{IC}^\text{int}_{\mu \times \mu}(\Pi_{f \times g}) = I(\Pi; X_1X_2|Y_1Y_2) + I(\Pi; Y_1Y_2|X_1X_2)
\]

\[
= \left[I(\Pi; X_1|Y_1Y_2) + I(\Pi; X_2|Y_1Y_2X_1)\right] + \left[I(\Pi; Y_2|X_1X_2) + I(\Pi; Y_1|X_1X_2Y_2)\right] \quad \text{(The chain rule)}
\]

\[
= \left[I(\Pi; X_1|Y_1Y_2) + I(\Pi; Y_1|X_1X_2Y_2)\right] \quad \text{(This will be} \geq \text{IC}^\text{int}_{\mu_f}(\Pi_f))
\]

\[
+ \left[I(\Pi; X_2|Y_1Y_2X_1) + I(\Pi; Y_2|X_1X_2)\right] \quad \text{(This will be} \geq \text{IC}^\text{int}_{\mu_g}(\Pi_g))
\]

We will finish the proof by constructing \( \Pi_f \) and \( \Pi_g \) whose int. inf. complexities are dominated by the last two terms in the above equation.
Protocol 1 The protocol $\Pi_f$

Given input $(x, y)$:

1. The parties use public randomness to sample $y_2 \sim Y_2$
2. Alice uses her private randomness to sample privately $x_2 \sim X_2|Y_2$
3. The parties simulate the protocol $\Pi_{f \times g}$ on $(x, y, x_2, y_2)$ and output accordingly

Protocol 2 The protocol $\Pi_g$

Given input $(x, y)$:

1. The parties use public randomness to sample $x_1 \sim X_1$
2. Bob uses his private randomness to sample privately $y_1 \sim Y_1|X_1$
3. The parties simulate the protocol $\Pi_{f \times g}$ on $(x_1, y_1, x, y)$ and output accordingly

Note that both $\Pi_f, \Pi_g$ compute $f, g$ (respectively) with $\epsilon$ error.

Consider $\Pi_f$ on input $(X_1, Y_1)$ which is drawn from $\mu_f$. Note that in this case, the simulation of $\Pi_{f \times g}$ by $f$ has the same distribution of $\Pi_{f \times g}$ on input drawn from $\mu_f \times \mu_g$. It follows that:

$$IC^\text{int}_{\mu_f} (\Pi_f) = I(\Pi_f; X_1|Y_1) + I(\Pi_f; Y_1|X_1)$$

$$= I(Y_2\Pi; X_1|Y_1) + I(Y_2\Pi; Y_1|X_1)$$

(The parties use public randomness to draw $y_2 \sim Y_2$)

$$= I(Y_2; X_1|Y_1) + I(\Pi; X_1|Y_1Y_2) + I(Y_2; Y_1|X_1) + I(\Pi; Y_1|X_1Y_2)$$

(chain rule)

$$= I(\Pi; X_1|Y_1Y_2) + I(\Pi; Y_1|X_1Y_2)$$

(from independence: $I(Y_2; X_1|Y_1) = I(Y_2; Y_1|X_1) = 0$)

$$\leq I(\Pi; X_1|Y_1Y_2) + I(\Pi; Y_1|X_1X_2Y_2)$$

$$= I(\Pi; Y_1|X_1Y_2)$$

The last inequality follows from a more general inequality for mutual information: Let $A, B, C, D$ be random variables such that $D$ is independent of $B$ given $C$. Then

$$I(A; B|C) \leq I(A; B|CD).$$

The proof of this inequality is left as an exercise (See Claim 1.1). Plugging in $A = \Pi, B = Y_1, C = X_1Y_2, D = X_2$ yields the last inequality.

In a similar way we show that

$$IC^\text{int}_{\mu_f} (\Pi_f) \leq I(\Pi; X_2|Y_1Y_2X_1) + I(\Pi; Y_2|X_1X_2).$$
This finishes the proof.

As a Corollary we obtain the following direct sum result for Internal Information:

**Theorem 4** (Direct Sum for Internal Information). Let \( f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z} \) be a function, let \( \mu \) be a distribution on \( \mathcal{X} \times \mathcal{Y} \) and let \( \epsilon \geq 0 \). Then:

\[
\IC_{\mu^{k},\epsilon}^{\text{int}} (f^k) \geq k \cdot \IC_{\mu,\epsilon}^{\text{int}} (f)
\]

We remark here that in the case of External Information, the above result do not hold for general \( \mu \). However, for full support \( \mu \) it is open.

**Exercise 2.2** (The Composition Lemma for Internal and External Information). Let \( \Pi_f \) be an \( \epsilon \)-error protocol for \( f \) and let \( \Pi_g \) be an \( \epsilon \)-error protocol for \( g \). Then there exists a marginal \( \epsilon \) error for \( f \times g \) such that for all \( \mu_f, \mu_g \):

\[
\IC_{\mu_f \times \mu_g}^{\text{int}} (\Pi_f \times g) = \IC_{\mu_f}^{\text{int}} (\Pi_f) + \IC_{\mu_g}^{\text{int}} (\Pi_g),
\]

and

\[
\IC_{\mu_f \times \mu_g}^{\text{ext}} (\Pi_f \times g) = \IC_{\mu_f}^{\text{ext}} (\Pi_f) + \IC_{\mu_g}^{\text{ext}} (\Pi_g)
\]

Conclude that if \( \epsilon = 0 \):

\[
\IC_{\mu^{n},\epsilon}^{\text{int}} (f^n) = n \cdot \IC_{\mu}^{\text{int}} (f)
\]

For analyzing the randomized communication complexity of Disjointness, we will need to following variant of The Composition Lemma for Internal Information:

**Lemma 2.3** (The \( \lor \)-Decomposition Lemma for Internal Information). Let \( f, g \) be boolean functions and let \( \Pi_{f \lor g} \) be an \( \epsilon \)-error protocol for \( f \lor g \), and let \( \mu_f \times \mu_g \) be a product distribution on \( \text{Dom}(f) \times \text{Dom}(g) \) such that \( \mu_f, \mu_g \) are supported on the zeros of \( f, g \). Then there exist \( \epsilon \)-error protocols \( \Pi_f, \Pi_g \) for \( f, g \) respectively such that:

\[
\IC_{\mu_f \times \mu_g}^{\text{int}} (\Pi_f \times g) = \IC_{\mu_f}^{\text{int}} (\Pi_f) + \IC_{\mu_g}^{\text{int}} (\Pi_g)
\]

Proving this Lemma is done in the same way we proved The Decomposition Lemma for Internal Information. The only subtle point one needs to consider is that since the distributions are supported by the zeros then the protocols \( \Pi_f, \Pi_g \) are indeed \( \epsilon \)-error protocols for \( f, g \).

As a Corollary we get

**Theorem 5** (Or-Lemma). Let \( f \) be a boolean function and let \( \mu \) be a distribution which is supported on the zeros of \( f \). Then:

\[
\IC_{\mu^n,\epsilon}^{\text{int}} (f^n) \geq n \cdot \IC_{\mu,\epsilon}^{\text{int}} (f)
\]
3 The randomized communication complexity of Disjointness

3.1 High level idea of the proof - reduction to internal information of AND

Let $f$ be a function. Assume we want to argue that $R_\epsilon(f) \geq T$. A simple observation is that it suffice to show that there exist a distribution $\mu$ for which every randomized $\epsilon$-error protocol $\Pi$ for $f$ has average communication complexity at least $T$:

$$CC^\text{avg}_\mu(\Pi) \geq T \implies \mathbb{E}_{x,y \sim \mu}[CC^\text{avg}_{x,y}(\Pi)] \geq T$$

$$\implies (\exists x,y) : CC^\text{avg}_{x,y}(\Pi) \geq T.$$ 

The second simple observation is:

$$\text{DISJ}(x,y) = \neg \bigvee_i x_i \land y_i.$$ 

Therefore, it is enough to consider the function $\bigvee_i x_i \land y_i$ = $\text{AND}^\vee n$. We will describe a particular $\mu$ which is supported on the zeros of $\text{AND}$ such that:

$$CC^\text{avg}_{\mu^n,\epsilon}(\text{AND}^\vee n) \geq IC^\text{int}_{\mu^n,\epsilon}(\text{AND}^\vee n) \geq n \cdot IC^\text{int}_{\mu,\epsilon}(\text{AND}) \geq \Omega(n).$$

By the Or-Lemma

In the next section we will finish the proof by giving a distribution $\mu$ which is supported on the zeros of $\text{AND}$ and for which

$$IC^\text{int}_{\mu,\epsilon}(\text{AND}) > 0.$$ 

In fact, we will give a large family of such $\mu$’s.

3.2 The Internal Information of AND

3.2.1 Divergence and Statistical Distance

Let $p, q$ be two distributions over the same sample space $X = \{1,\ldots,n\}$. The Divergence of $p$ relative to $q$ has the following intuition:

“Originally we thought that $X$ is distributed according to $q$, we learnt that $X$ is distributed according to $p$. The divergence of $p$ relative to $q$ is the amount of information we learnt.”
It is defined by:
\[
\mathbb{D}(p||q) = \sum_i p_i \log \frac{p_i}{q_i}.
\]

Remarks:

- \(\mathbb{D}(p||q)\) is always greater or equal than 0. Moreover, equality holds if and only if \(p = q\). Proving this is left as an exercise (Hint: Jensen’s Inequality)

- If there exist \(i\) such that \(q_i \neq 0 \neq p_i\) then \(\mathbb{D}(p||q) = \infty\). This somehow suits the intuition since if we thought that \(X\) is distributed according to \(q\) then in particular we were absolutely certain that the event \(i\) never happens. Learning that this event is possible to happen gave us infinite information. (If you are still not convinced, imagine you are a mathematician: mathematicians say “never” only when they are able to prove it. Finding a mistake in a proof of a mathematician gives him infinite information :) )

- Another argument which supports this intuition relies on the following equality:

\[
\mathbb{D}(p||q) = \sum_i p_i \log p_i + \sum_i p_i \log q_i = \sum_i p_i \log \frac{1}{q_i} - \sum_i p_i \log \frac{1}{p_i}.
\]

Note that \(\sum_i p_i \log \frac{1}{q_i}\) is (approximately) the average length if we would encode \(X \sim p\) with a Huffman code assuming that \(X \sim q\) and \(\sum_i p_i \log \frac{1}{p_i}\) is (approximately) the average length if we would encode \(X \sim p\) after learning that \(X \sim p\). Thus, \(\mathbb{D}(p||q)\) is the average number of bits we saved due to learning that \(X \sim p\)

- **Divergence and metrics**: It is very tempting to think of the divergence as if it satisfies the properties of a metric or at least that it satisfies the triangle inequality. Confusingly, it is not the case. However, the following Inequality by Pinsker, shows that if \(p, q\) are far from each other then also \(\mathbb{D}(p||q)\) is big:

\[
\mathbb{D}(p||q) \geq \frac{1}{2 \ln 2} (||p, q||_1)^2
\]

- **Divergence and Mutual Information**: The intuition we gave for Mutual Information might cause some confusion between the two quantities. However, the following
relation between the two settles it:

\[ I(X, Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x) \cdot p(y)} \]

\[ = \sum_x p(x) \sum_y p(y|x) \log \frac{p(y|x)}{p(y)} \]

\[ = \sum_x p(x) D(p(Y|x) || p(Y)) \]

Thus, the mutual information between \( X, Y \) is the average over \( X \) of the information we gained if we learned that \( X = x \).

### 3.2.2 High level idea of the proof

Let \( \mu \) be the distribution:

\[
\begin{array}{c|c|c}
 x \setminus y & y=0 & y=1 \\
\hline
 x=0 & \frac{1}{3} & \frac{1}{3} \\
 x=1 & \frac{1}{3} & 0 \\
\end{array}
\]

Let \( \Pi \) be a randomized \( \epsilon \)-error protocol for \( AND \). For \( x, y \in \{0, 1\} \) let \( \Pi_{x,y} \) denote the distribution of \( \Pi(x,y) \). The crux of the proof is to find two distributions \( \Pi_{x',y'} \) and \( \Pi_{x'',y''} \) which are “far away” from each other such that both \( (x', y') \) and \( (x'', y'') \) are in the support of \( \mu \). We will see that choosing \( (1, 0) \) and \( (0, 1) \) will do the job. If one thinks about it a bit one sees it is quite intuitive since this pair of inputs is a fooling pair and therefore, due to the rectangle property, it is likely to expect that \( \Pi_{0,1} \) and \( \Pi_{1,0} \) are very different distributions.

Assume we know that \( \Pi_{0,1} \) and \( \Pi_{1,0} \) are statistically far away from each other. How do we get that \( IC_{\mu}^{\text{int}}(\Pi) \) is \( \Omega(1) \)? The idea is to express the internal information as a sum of divergences between distribution of the form \( \Pi_{x,y} \) and then to use that if \( ||\Pi_{0,1}, \Pi_{0,0}|| \) is \( \Omega(1) \) then by Pinsker Inequality also the internal information is \( \Omega(1) \):

\[
IC_{\mu}^{\text{int}}(\Pi) = I(\Pi; X|Y) + I(\Pi; Y|X)
\]

\[
= \Pr(Y = 0)I(\Pi; X|Y = 0) + \Pr(Y = 1)I(\Pi; X|Y = 1)
\]

\[
+ \Pr(X = 0)I(\Pi; Y|X = 0) + \Pr(X = 1)I(\Pi; Y|X = 1)
\]

\[
= \Pr(Y = 0)I(\Pi; X|Y = 0) + \Pr(X = 0)I(\Pi; Y|X = 0) .
\]

(since \( I(\Pi; X|Y = 1) = I(\Pi; Y|X = 1) = 0 \))
Now, let us express $I(\Pi; X|Y = 0)$ in terms of the the Divergence:

$$I(\Pi; X|Y = 0) = \Pr(X = 1|Y = 0) \mathcal{D}(\Pi_{1,0}||\Pi_{Y=0}) + \Pr(X = 0|Y = 0) \mathcal{D}(\Pi_{0,0}||\Pi_{Y=0})$$

$$= \frac{1}{2} \mathcal{D}(\Pi_{1,0}||\Pi_{Y=0}) + \frac{1}{2} \mathcal{D}(\Pi_{0,0}||\Pi_{Y=0})$$

$$\geq \frac{1}{4\ln 2} (||\Pi_{1,0},\Pi_{Y=0}||^2 + ||\Pi_{0,0},\Pi_{Y=0}||^2) \quad \text{(By Pinsker Inequality)}$$

$$\geq \frac{1}{2\ln 2} (||\Pi_{1,0},\Pi_{Y=0}|| + ||\Pi_{0,0},\Pi_{Y=0}||)^2 \quad \text{(By convexity of } x \mapsto x^2)$$

$$\geq \frac{1}{2\ln 2} (||\Pi_{1,0},\Pi_{0,0}||)^2 \quad \text{(By triangle Inequality)}$$

Similarly, we show that:

$$I(\Pi; Y|X = 0) \geq \frac{1}{2\ln 2} (||\Pi_{0,1},\Pi_{0,0}||)^2.$$

Thus we get that:

$$IC_{\mu}^{\text{int}}(\Pi) \geq \frac{1}{2\ln 2} (||\Pi_{0,1},\Pi_{0,0}||)^2.$$

By another application of convexity of $x^2$ and the triangle inequality, we get that

$$IC_{\mu}^{\text{int}}(\Pi) \geq \frac{1}{\ln 2} (||\Pi_{0,1},\Pi_{1,0}||)^2.$$

Recall that a pair of inputs $(x', y'), (x'', y'')$ is called a fooling pair w.r.t $f$ if there exist no $f$-monochromatic rectangle which contains the two inputs. A special property of the pair of inputs $(0, 1)$ and $(1, 0)$ w.r.t $\text{AND}$ is that it is a fooling pair. The final step of the proof is to show that the right hand side of the above inequality is $\Omega(1)$. The following Lemma, which will be proven in the next section implies it:

**Lemma 3.1** (Fooling pair for randomized protocols). Let $f$ be a function, let $(x', y'), (x'', y'')$ be a fooling pair and let $\Pi$ be an $\epsilon$-error private coin protocol for $f$. Then:

$$||\Pi_{x',y'}',\Pi_{x'',y''}'||_1 \geq 1 - 3\epsilon$$

**3.2.3 Fooling sets for randomized protocols**

**Lemma 3.2** (The rectangle property for private coin protocols). Let $\Pi$ be a private coin protocols and let $V$ denote the set of leaves of $\Pi$. There exist functions $a : V \times X \rightarrow [0, 1]$, $b : V \times Y \rightarrow [0, 1]$ such that for every $v \in V$:

$$\Pr(\Pi(x, y) \text{ reaches } v) = a(v, x) \cdot b(v, y).$$
Proof. Let \( v_0, v_1, ..., v_k = v \) be the vertices on the path from the root to \( v \). Then:

\[
\Pr (\Pi (x, y) \text{ reaches } v) = \Pr (v_1|v_0, x, y) \cdot \Pr (v_2|v_1, x, y) \cdots \Pr (v_k|v_{k-1}, x, y).
\]

Let \( I_a = \{i|v_i \text{ is owned by Alice}\} \) and let \( I_b = \{i|v_i \text{ is owned by Bob}\} \). Define:

\[
a(v, x) = \prod_{i \in I_a} \Pr (v_{i+1}|v_i, x, y) = \prod_{i \in I_a} \Pr (v_{i+1}|v_i) \quad (\text{The } v_i \text{'s are owned by Alice - they do not depend on } y.)
\]

Similarly, define

\[
b(v, y) = \prod_{i \in I_b} \Pr (v_{i+1}|v_i, y).
\]

Thus, we have

\[
\Pr (\Pi (x, y) \text{ reaches } v) = a(v, x) \cdot b(v, y).
\]

We are now ready to prove The fooling pairs for randomized protocols Lemma (Lemma 3.1):

**Proof.** We distinguish between two types of fooling pairs \( \{(x', y'), (x'', y'')\} \):

1. \( f(x', y') \neq f(x'', y'') \)
2. \( f(x', y') = f(x'', y'') \neq f((x', y'') \)

For the first type, the correctness of \( \Pi \) implies that \(|\Pi_{x',y'}, \Pi_{x'',y''}|_1 \geq 1 - 2\epsilon. \) (Why?)

Consider a fooling pair \( \{(x', y'), (x'', y'')\} \) of the second type. Assume w.l.o.g that \( f(x', y') = f(x'', y'') = 1. \) Let \( \mathcal{P} \) be the set of all leaves of \( \Pi \) which are labeled by 1. From the correctness of \( \Pi \), it follows that:

1. \( f(x', y') = 1 \implies \sum_{v \in \mathcal{P}} a(v, x') \cdot b(v, y') = \Pr (\Pi (x', y') = 1) \geq 1 - \epsilon \)

2. \( f(x'', y'') = 1 \implies \sum_{v \in \mathcal{P}} a(v, x'') \cdot b(v, y'') = \Pr (\Pi (x'', y'') = 1) \geq 1 - \epsilon \)

3. \( f(x'', y') \neq 1 \implies \sum_{v \in \mathcal{P}} a(v, x') \cdot b(v, y'') = \Pr (\Pi (x', y'') = 1) \leq \epsilon \)
4.\[ \sum_{v \in P} a(v, x'') \cdot b(v, y') = \Pr (\Pi(x'', y') = 1) \leq 1 \]

Therefore:
\[ \|\| \Pi(x', y') , \Pi(x'', y'') \|\| = \sum_{v \in \Pi} |a(v, x') b(v, y') - a(a, x'') b(b, y'')| \]
\[ \geq \sum_{v \in P} |a(v, x') b(v, y') - a(a, x'') b(b, y'')| \]
\[ \geq \sum_{v \in P} \left( \sqrt{a(v, x') b(v, y')} - \sqrt{a(a, x'') b(b, y'')} \right)^2 \]
\[ (\forall t, s) : |t - s| \geq (\sqrt{t} - \sqrt{s})^2 \]
\[ = \sum_{v \in P} a(v, x') b(v, y') + \sum_{v \in P} a(v, x'') b(v, y'') \]
\[ - \sum_{v \in P} 2\sqrt{a(v, x') b(v, y') a(v, x'') b(v, y'')} \]
\[ \geq 2 - 2\epsilon - \sum_{v \in P} 2\sqrt{a(v, x') b(v, y') a(v, x'') b(v, y'')} \]
\[ \text{(from the 1st and 2nd bullets above)} \]
\[ \geq 2 - 2\epsilon - \sum_{v \in P} a(v, x') \cdot b(v, y'') + a(v, x'') \cdot b(v, y') \]
\[ \text{(The Arithmetic-Geometric means inequality)} \]
\[ \geq 2 - 2\epsilon - (\epsilon + 1) \]
\[ \text{(from the 3rd and 4th bullets above)} \]
\[ = 1 - 3\epsilon \]

Exercise 3.3. Let $f$ be a function, let $(x', y'), (x'', y'')$ be a fooling pair and let $\Pi$ be an 0-error private coin protocol for $f$. Then:
\[ \|\| \Pi_{x', y'}, \Pi_{x'', y''} \|_1 = 2. \]

The following exercise is harder than the rest of the exercises.

Exercise 3.4 (Pairwise far probability vectors). Let $v_1, ..., v_k$ be probability vectors in $\mathbb{R}^n$ and let $d > 0$ be a constant such that
\[ (\forall i \neq j) : \|v_i, v_j\|_1 \geq d. \]

Show that:
- If $d = 2$ then $k \leq n$
• If \( d < 2 \) then \( k = \mathcal{O}(2^n) \)

Conclude that for any function \( f \) with a fooling set of size \( k \):

• \( R_0(f) \geq \log k \)

• \( R(f) = \Omega(\log \log k) \)

**Exercise 3.5** (Communication Complexity is useful in geometry!). Let \( n \in \mathbb{N} \). Give an explicit construction of probability vectors \( v_1, \ldots, v_k \) such that

• \((\forall i \neq j) : \|v_i, v_j\|_1 \geq 1.\)

• \( k = \Omega(2^n) \)

**Hint:** Think of a function \( f \) with efficient randomized protocol and a large fooling set.

### 3.3 A short summary of the proof and some remarks

The above proof relied heavily on the direct sum of internal information. Applying the direct sum enabled us to reduce the analysis of the randomized communication complexity of \( \text{DISJ}_n \) for **every** \( n \) to analyzing the internal information of the \( \text{AND} \) function which is **one particular** 2-bit function. An important relation we used is the following: If \( f \) is a boolean function and \( \mu \) is a distribution which is supported on the zeros of \( f \) then:

\[
\text{IC}_{\mu}^{\text{int}}(f, \epsilon) > 0 \implies \text{IC}_{\mu^n, \epsilon}^{\text{int}}(f^\lor n) = \Omega(n).
\]

It is easy to verify that the proof we gave to this relation actually proves something stronger: Instead of considering \( f^\lor n \), one can consider any function \( g \) which agrees with \( f^\lor n \) on inputs for which there is at most one \( i \) such that \( f(x_i, y_i) = 1 \). Thus, we actually proved the stronger variant of the hardness of set-disjointness which we discussed in class.