15-859: Information Theory and Applications in TCS **CMU: Spring 2013**

Lecture 21: Set Disjointness lower bound via product distribution

April 11, 2013

Lecturer: Venkatesan Guruswami *Scribe: Shashank Singh*

1 Recap

Last lecture we covered the following:

- Showed $R(IP) = \Theta(n)$ using the Discrepancy Method
- Indexing Problem: showed Alice must sent $\geq \Omega(n)$ bits using Information Theory

2 Set Disjointness lower bound via product distribution

Today we lower bound $R(\text{DISJ})$, where

$$\text{DISJ}(x, y) = \bigwedge_{i=1}^{n} \text{NAND}(x_i, y_i).$$

2.1 Preliminary Observations

Our goal is choose μ such that $D^{\mu}_{1/100}(\text{DISJ})$ is large. Notice that if, for example, μ is uniform, then the probability that $\text{DISJ}(x, y)) = 1$ is $(3/4)^n$, and so Alice and Bob can correctly guess "not disjoint" with high probability.

Thus, μ should be "balanced" in the sense that

$$\mu(\text{DISJ}^{-1}(0)), \mu(\text{DISJ}^{-1}(1)) = \Omega(1).$$

Remark 1 *Consider μ with $x_1, \dots, x_n, y_1, \dots, y_n \sim$ i.i.d. Bernoulli$(1/\sqrt{n})$. This μ is "balanced", since*

$$\lim_{n\to\infty} \mathsf{P}(\text{DISJ}(x, y) = 1) = \lim_{n\to\infty} (1 - \mathsf{P}(x_i \wedge y_i))^n = \lim_{n\to\infty} \left(1 - \frac{1}{n}\right)^n = 1/e.$$

Proposition 2 *(Babai, Frankl, Simon, 1986) Consider μ with $x_1, \dots, x_n, y_1, \dots, y_n \sim$ i.i.d. Bernoulli$(1/\sqrt{n})$. Then, $D^{\mu}_{1/100}(\text{DISJ}) = \Omega(\sqrt{n})$ (in fact, $D^{\mu}_{1/100}(\text{DISJ}) = \Theta(\sqrt{n})$).*

Corollary 3 $R(\text{DISJ}) \geq \Omega(\sqrt{n})$.

2.2 Proof of Proposition 2

Suppose Π_0 is a deterministic protocol such that

$$\mathop{\mathsf{P}}_{(x,y)\sim\mu}(\mathrm{DISJ}(x,y)=\Pi_0(x,y)) \geq 0.99.$$

Let the random variable Π denote the transcript (log of bits sent) of Π_0 on $(x,y)\sim\mu$. We know

$$\begin{aligned} CC(\Pi_0) &\geq \log_2 \big|\operatorname{supp}(\Pi)\big| \\ &\geq H(\Pi(X,Y)) = I(X,Y\,;\,\Pi) \\ &= I(x_1,\dots,x_n,y_1,\dots,y_n\,;\,\Pi) \\ &\geq \sum_{i=1}^{n} I(x_i,y_i\,;\,\Pi). \end{aligned}$$

Definition 4

$$\Pi^i_{a,b} \triangleq \Pi \text{ conditioned on } x_i=a, y_i=b.$$

In Problem 6 of Problem Set 1, we showed

$$I(x_i,y_i\,;\,\Pi) \geq \mathop{\mathbb{E}}_{(a,b)\sim(\mathrm{Ber}(1/\sqrt{n}))^2}\left[\Delta^2_{TV}\left(\Pi^i_{a,b},\Pi\right)\right],$$

where

$$\Delta_{TV}(A,B) \triangleq \frac{1}{2}\sum_{\ell}\big|\,\mathsf{P}(A=\ell)-\mathsf{P}(B=\ell)\big|.$$

Thus, noting $\frac{1}{\sqrt{n}}\left(1-\frac{1}{\sqrt{n}}\right) \geq \frac{1}{2\sqrt{n}}$,

$$\begin{aligned} I(x_i,y_i\,;\,\Pi) &\geq \frac{1}{\sqrt{n}}\left(1-\frac{1}{\sqrt{n}}\right)\left[\Delta^2_{TV}\left(\Pi^i_{1,0},\Pi\right)+\Delta^2_{TV}\left(\Pi^i_{0,1},\Pi\right)\right] \\ &\geq \frac{1}{4\sqrt{n}}\left[\Delta_{TV}\left(\Pi^i_{1,0},\Pi\right)+\Delta_{TV}\left(\Pi^i_{0,1},\Pi\right)\right]^2 \\ &\geq \frac{1}{4\sqrt{n}}\left[\Delta_{TV}\left(\Pi^i_{1,0},\Pi^i_{0,1}\right)\right]^2, \end{aligned}$$

where the last inequality is by the Triangle Inequality, since Δ_{TV} is a metric. Thus, we've shown so far that

$$\begin{aligned} CC(\Pi_0) &\geq n\,\mathop{\mathbb{E}}_{i}\left[I(x_i,y_i;\Pi)\right] \\ &\geq \frac{n}{4\sqrt{n}}\mathop{\mathbb{E}}_{i}\left[\Delta^2_{TV}\left(\Pi^i_{1,0},\Pi^i_{0,1}\right)\right] \\ &\geq \frac{\sqrt{n}}{4}\mathop{\mathbb{E}}_{i}\left[\Delta_{TV}\left(\Pi^i_{1,0},\Pi^i_{0,1}\right)\right]^2. \end{aligned}$$

Now, it suffices to show that

$$\mathop{\mathbb{E}}_{i}\left[\Delta_{TV}\left(\Pi^i_{1,0},\Pi^i_{0,1}\right)\right]^2 \geq \Omega(1).$$

We break the proof of this into two lemmas:

Lemma 5 *Since* Π_0 *computes* DISJ *with high accuracy,*

$$\mathop{\mathbb{E}}_{i}\left[\Delta_{TV}\left(\Pi^i_{0,0},\Pi^i_{1,1}\right)\right]=\Omega(1).$$

Lemma 6 *If* $\Delta_{TV}\left(\Pi^i_{0,0},\Pi^i_{1,1}\right)\geq\Omega(1)$*, then* $\Delta_{TV}\left(\Pi^i_{0,1},\Pi^i_{1,0}\right)\geq\Omega(1)$.

Proof: (of Lemma 5) Since $\mathsf{P}\left(\mathrm{DISJ}(X,Y)=1\,|\,X_i=Y_i=0\right)\geq 1/4$,

$$\mathsf{P}\left(\Pi_0(\Pi^i_{0,0})=1\right)\geq 1/5,$$

where $\Pi_0(\Pi^i_{0,0})$ is the output of Π_0 given the transcript $\Pi^i_{0,0}$. Since $X_i=Y_i=1\Rightarrow\mathrm{DISJ}(X,Y)=0$,

$$\mathsf{P}\left(\Pi_0(\Pi^i_{1,1})=1\right)\leq 1/6.$$

Thus,

$$\Delta_{TV}(\Pi^i_{0,0},\Pi^i_{1,1})\geq 1/5-1/6=1/30.$$

Hence, Π_0 is, on average, a good distinguisher of $\Pi^i_{0,0}$ and $\Pi^i_{1,1}$. ∎

Proof: (of Lemma 6) We make use of the Hellinger distance:

Definition 7 *The Hellinger distance between two random variables A and B is*

$$\Delta_{\text{Hel}} \triangleq \sqrt{1 - \sum_{\ell} \sqrt{\mathsf{P}(A=\ell)\,\mathsf{P}(B=\ell)}} = \sqrt{1 - Z(A,B)},$$

where $Z(A,B)$ denotes the Bhattacharya coefficient.

Exercise Use Cauchy-Schwarz to show

$$\Delta^2_{\text{Hel}}(A,B) \le \Delta_{TV}(A,B) \le \sqrt{2}\Delta_{\text{Hel}}(A,B).$$

By this Exercise, it suffices to show that

$$\Delta^2_{\text{Hel}}(\Pi^i_{0,0}, \Pi^i_{1,1}) = \Delta^2_{\text{Hel}}(\Pi^i_{0,0}, \Pi^i_{1,1}),$$

and hence it suffices to show, for each i,

$$\mathsf{P}\left(\Pi^i_{0,0} = \tau\right) \mathsf{P}\left(\Pi^i_{1,1} = \tau\right) = \mathsf{P}\left(\Pi^i_{0,1} = \tau\right) \mathsf{P}\left(\Pi^i_{1,0} = \tau\right).$$

Fix i and recall the following Rectangle Property:

- *Inputs $X^{-i} := (X_1, \ldots, X_{i-1}, X_{i+1}, X_n), Y^{-i} := (Y_1, \ldots, Y_{i-1}, Y_{i+1}, Y_n)$ leading to a transcript τ form a rectangle $R_\tau = S_\tau \times T_\tau$. Since $X \perp Y$,*

$$\mathsf{P}\left(\Pi^i_{a,b} = \tau\right) = \mathsf{P}\left(X^{-i} \in S_\tau \wedge Y^{-i} \in T_\tau\right) = \mathsf{P}\left(X^{-i} \in S_\tau\right) \mathsf{P}\left(Y^{-i} \in T_\tau\right) = A_a(\tau) B_b(\tau).$$

Importantly, $\mathsf{P}\left(\Pi^i_{a,b} = \tau\right)$ factors into non-negative functions A_0, A_1, B_0, B_1. Thus,

$$\begin{aligned}
\mathsf{P}\left(\Pi^i_{0,0} = \tau\right) \mathsf{P}\left(\Pi^i_{1,1} = \tau\right) &= A_0(\tau) B_0(\tau) A_1(\tau) B_1(\tau) \\
&= A_0(\tau) B_1(\tau) A_1(\tau) B_0(\tau) \\
&= \mathsf{P}\left(\Pi^i_{0,1} = \tau\right) \mathsf{P}\left(\Pi^i_{1,0} = \tau\right).
\end{aligned}$$

■

Remark 8 *Babai, Frankl, and Simon (1986) also showed that, for any μ which can be factored as a product distribution (meaning $\mu(x,y) = \mu_A(x) \cdot \mu_B(y)$),*

$$D^\mu(\text{DISJ}) = O(\sqrt{n} \log n).$$

Thus, getting a substantially better lower bound requires adding correlation between X and Y.

3 Next Time

Next time, we will show $R(\text{DISJ}) = \Omega(n)$.

- This result was first shown by Kalyanasundaram and Schnitger (1987).
- Razborov (1990) "simplified" the proof.
- We'll see an Information Theory based proof by Bar-Yossef, Jayram, Kumar, Sivakumar (2004).