CS 229r Information Theory in Computer Science	April 9th, 2019
Lecture 19	
Instructor: Madhu Sudan	*Scribe: Stefan Spataru*

1 Overview

Today we'll prove the parallel repetition theorem. The references are

1. Ran Raz 94 [1]
2. Holenstein: Simplified proof. [2]
3. Barak: Notes '09 [3].

[1] is the original paper, [2] provides a shorter proof, while the lecture mostly follows the exposition in [3]

2 2-prover-Game

Definition 1. *A 2-**prover game** is a 6-tupe $G = (X, Y, A, B, \mu, V)$, where μ is a distribution on $X \times Y$, and V is a verification function $V : X \times Y \times A \times B \to \{0, 1\}$. A strategy is $f : X \to A$ and $g : Y \to B$. The value of the strategy*

$$val_{f,g}(G) = \mathbf{E}_{x,y\sim\mu} V(x, y, f(x), g(y))$$

Definition 2. *One will define the k-fold repetition of a game G as*

$$G^{\otimes k} = (X^k, Y^k, A^k, B^k, \mu^k, V^k)$$

where $V^k(\overline{x}, \overline{y}, \overline{a}, \overline{b}) = \bigwedge_{i=1}^{n} V(x_i, y_i, a_i, b_i)$.

3 Parallel repetition theorem

Theorem 3. *(Ran Raz)*
$\forall \varepsilon > 0, A, B,$ exists $\delta > 0$ such that for all G, k

$$w(G) \leq 1 - \varepsilon \Rightarrow w(G^{\otimes k}) \leq (1 - \delta)^k$$

For the purposes of proving this lemma, define:

Definition 4. *If one fixes strategy f, g, one defines W_S to be the event $\bigwedge_{i\in S} V(x_i, y_i, f(x)_i, g(x)_i)$.*

We aim to prove the following lemma:

Lemma 5. *For any $\varepsilon > 0$, A, B, there exists $\gamma > 0$ such that if $w(G) \geq 1 - \varepsilon$ and*
$\mathbb{P}(W_s) \leq 2^{-\gamma k}$ or there exists $i \notin S$ $\mathbb{P}(W_{\{i\}} \mid W_s) \leq 1 - \frac{\varepsilon}{100}$

Remark

The parallel repetition theorem follows immediately from the lemma. As such, one keeps adding indices until the probability becomes low enough. If all the indices have been added, then a contradiction is reached, as at each step probability decreased by a factor of $1 - \frac{\varepsilon}{100}$.

Example 6. *Let $X = \{1, 2\}$, $Y = \{3, 4\}$, $\mu \sim Unif(X \times Y)$, $A = B = \{1, 2, 3, 4\}$. $V(x, y, a, b) = 1$ if $a = b$ and $b \in \{x, y\}$. Now, $w(G) = w(G^{\otimes 2})$ as we proved in the last class. Also, $w(G^{\otimes k}) \geq 2^{-\frac{k}{2}}$ by pairing coordinates.*

3.1 Reduction

3.1.1 Intuition

We will first provide some intuition for the algorithm:

The proof will work by reduction. Take a copy of the game **Alice** $\leftarrow U$ and $V \rightarrow$ **Bob**.r will be assumed to be small compared to k. (since one in interested in small γ). We will be interested in acceptance in the k-fold version of this game. For simplicity, we will assume WLOG $S = \{1, ..., r-1\}$. We are planning to inject U into coordinate r into X and V into coordinate r into Y. The argument needs one to pick $X_{<r}$ and $Y_{<r}$ so that we win on coordinates $s < r$. Choosing these coordinates creates conditioning on X_r, Y_r. Thus, we will be forced to condition on questions/answers. Turns out we will only be forced to conditions on answers on $s < r$. The strategy here will be based off the fact that one is interested in small γ, and so $r = \gamma k$ is small compared to k.

Next, one samples $X_{>r}$, $Y_{>r}$. The problem for sampling now is that there are too many coordinates. For practical purposes $\gamma k \approx k$. The strategy here will be following: We need a large portion of the space, as we know there are many points where we win the game. Conditioning on this would be too much as there are very few coordinates left. The solution comes from the fact that we don't have to win on coordinates $k > r$. So all we need is X_r, Y_r to be from the correct distribution roughly, i.e. μ^{k-r} (even conditioned on the first r coordinates), but we don't have to condition on $a_{>r}, b_{>r}$.

We would like to come up with $X_1, ..., X_{k-1}, Y_1, ..., Y_{k-1}$. We fixed f, g. We want $X_{<r}, Y_{<r}$ such that $V(X+j, Y_j, f(X^k)_j, g(X^k)_j) = 1$ and $\mathbb{P}(V(X_r, Y_r, f(X^k)_r, g(X^k)_r)) \geq 1 - \frac{\varepsilon}{2}$. If one wants such a strategy, let $T_{>r}$ be the common randomness. $T_j = (0/1, x_j/y_j)$. The first coordinate tells whether we are sampling over the marginal distribution of x or y. Here's how Alice will compute $X_{>r}$ :: Given $T_{>r}$, $X_j = x_j$ if $(T_j) = (0, x_j)$, and $X_j = X|_{Y_j = y_j}$ if $t_j = (1, y_j)$. If this wins the first $r-1$ coordinates accept, otherwise repeat. There is still randomness after T_j. The motivation for using T_j is that it gives common randomness. Fix also answers on the first $r-1$ coordinates. Alice assumes answers are $a_1, ..., a_{r-1}$. Bob assumes his answers are $b_1, ..., b_{r-1}$.

Up until now we assumed looking at one r. But there is no reason to. One only needs to pick $i \in \{r, ..., k\}$ such that conditional probability is small. What makes i good enough? Luck or careful examination? What should this i satisfy? Pick "typical" $X_{<r-1}, Y_{<r-1}, a_{<r-1}, b_{<r-1}$ (specify later). Let W be the event that the first $r-1$ questions are $X_{<r}$ and that they are $Y_{<r}$ and the first $r-1$ answers are $a_{<r}$, $b_{<r}$ respectively when $X_{<r}$, $Y_{<r}, a_{<r}, b_r$ are winning.

1. The first condition we would like is $(X_i, Y_i)|W \approx (U, V)$.

2. The second condition refers to $T_{>r} \;|_{W, X_i, Y_i}$. Neither Bob nor Alica can sample from this as they do not have access to X_i, Y_i respectively. The second condition we want is that $T_{>r}|_{W, X_r, Y_r} \approx T_{>r}|_{W, X_r} \approx T_{>r}|_{W, Y_r}$. We can now use correlated sampling.

Alice sets $X_1, ..., X_{r-1}$ as above. Next, she sample the remaining coordinates. First, she samples $X_{1,...,r-1}|W, X_{<r}$. Correlate this with $T_{>r}|W, Y_r$ which is what Bob will do. Having done this, they will complete $X_{r+1,...,X_n}|W$ and $Y_{r+1}, ..., Y_r|W$. Resample if necessary. The claim is that probability of choosing a "good" i is high.

3.2 The reduction formally:

We summarize the reduction below:

4 Aside

How do we prove a statement of the form the distributions "look the same"? Mostly through the following lemma:

Algorithm 1 Reduction

Pick $x_{<r}, y_{<r}, a_{<r}, b_{<r}$ typically.
Set $X_{<r} = x_{<r}$, $Y_{<r} = y_{<r}$.
Pick the correlators T_j defined above.
Let W be the event that $f(x_{<r}) = a_{<r}$, $g(y_{<r}) = b_{<r}$ and $W_{<r}$.
Alice computes $X_{>r}$ from T_r as described above.
Compute $f(X^k)_r, f(Y^k)_r$.

Lemma 7. *If $(X_1, Y_1), ..., (X_n, Y_n)$ are independent and E is some event of probability $\geq 2^{-d}$, then*

$$D((X^n, Y^n) \mid E||(X^n, Y^n)) \leq d$$

Using independence,

$$\mathbf{E}_{i \sim Unif} D((X_i, Y_i) \mid E||(X_i, Y_i)) \leq \frac{d}{n}$$

and so

$$\mathbf{E}_i(\delta((X_i, Y_i)|E, (X_i, Y_i)) \leq \mathcal{O}(\sqrt{\frac{d}{n}})$$

References

[1] Raz, R. (1998). A Parallel Repetition Theorem. SIAM Journal on Computing, 27(3), 763-803

[2] Holenstein, T. (2007). Parallel repetition: Simplifications and the no-signaling case. Proceedings of the Thirty-ninth Annual ACM Symposium on Theory of Computing, 411-419.

[3] Barak, B. (2009) http://www.cs.princeton.edu/courses/archive/spr07/cos522/ho11.pdf