1 Overview

Today we’ll prove the parallel repetition theorem. The references are
1. Ran Raz 94 [1]

[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-tuple $G = (X, Y, A, B, \mu, V)$, where $\mu$ 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 $val_{f,g}(G) = \mathbb{E}_{x,y \sim \mu} V(x, y, f(x), g(y))$.

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

3 Parallel repetition theorem

Theorem 3. (Ran Raz)

\[ \forall \epsilon > 0, A, B, \text{ exists } \delta > 0 \text{ such that for all } G, k \]

\[ w(G) \leq 1 - \epsilon \Rightarrow w(G^{\otimes k}) \leq (1 - \delta)^k \]

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 $\epsilon > 0, A, B, \text{ there exists } \gamma > 0 \text{ such that if } w(G) \geq 1 - \epsilon$ and

\[ \mathbb{P}(W_s) \leq 2^{-\gamma k} \text{ or there exists } i \not\in S \mathbb{P}(W_{(i)} | W_s) \leq 1 - \frac{\epsilon}{100} \]

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 $1 - \frac{\epsilon}{100}$.

Example 6. Let $X = \{1, 2\}, Y = \{3, 4\}, \mu \sim \text{Unif}(X \times Y), A = B = \{1, 2, 3, 4\}$. $V(x, y, a, b) = 1$ if $a = b$ and $b \in \{x, y\}$. Now, $w(G) = w(G^{\otimes 2})$ as we proved in the last class. Also, $w(G^{\otimes k}) \geq 2^{\frac{k}{10}}$ 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 ← U and V → Bob, r will be assumed to be small compared to k. (since one is 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<, and Y< so that we win on coordinates s < r. Choosing these coordinates creates conditioning on Xr,Yr. 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 = γk 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 γk ≈ 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,Y to be from the correct distribution roughly, i.e. µr−k (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 X1,...,Xk−1, Y1,...,Yk−1. We fixed f,g. We want X<, Y< such that V(X + jY, f(Xk))j,g(Xk)j) = 1 and P(V(Xr,Yr,f(Xk)r,g(Xk)r)) ≥ 1 − 2 Γ. If one wants such a strategy, let T>_r be the common randomness. T j = (0/1,xj/yj). 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, Xj = xj if (T j) = (0,xj), and Xj = X|Yj=yj if t j = (1,yj). 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 a1,...,ar−1. Bob assumes his answers are b1,...,br−1.

Up until now we assumed looking at one r. But there is no reason to. One only needs to pick i ∈ {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 (Xj,Yj)|W ≈ (U,V).

2. The second condition refers to T>_r |W,X,Y. Neither Bob nor Alice can sample from this as they do not have access to X1,Y respectively. The second condition we want is that T>_r|W,X,Y ≈ T>_r|W,Y. We can now use correlated sampling.

Alice sets X1,...,Xr−1 as above. Next, she sample the remaining coordinates. First, she samples X1,...,r−1|W,X<_r. Correlate this with T>_r|W,Y, which is what Bob will do. Having done this, they will complete Xr+1,...,X|W and Yr+1,...,Y|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

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

Using independence,

$$E_{i \sim \text{Unif}} D((X_i, Y_i) \mid E \parallel (X_i, Y_i)) \leq \frac{d}{n}$$

and so

$$E_i(\delta((X_i, Y_i) \mid E, (X_i, Y_i)) \leq O(\sqrt{\frac{d}{n}})$$

References

