COS 522: Complexity Theory : Boaz Barak
Handout 10: Parallel Repetition Lemma

Reading: **(1)** A Parallel Repetition Theorem / Ran Raz (available on his website) **(2)** Parallel Repetition: Simplifications and the No-Signalling Case / Thomas Holenstein `http://arxiv.org/abs/cs/0607139`

Parallel Repetition We state and prove it for 2 prover games, can be easily generalized to 2-query PCP's (see Exercise 1).

Def: In a two-prover game:

1. Two provers first coordinate two strategies P_1 and P_2 (possibly with shared randomness).
2. The verifier selects a random pair of strings (x, y) from a joint distribution (X, Y) and gives each input to each prover respectively.
3. We denote by $a = P_1(x)$ the first prover's answer and $b = P_2(y)$ the second prover's answer.
4. The verifier decides to accept or not based on some predicate $V(c, y, a, b)$.

The *value* of the game is the maximum over all strategies P_1, P_2 of the probability that V accepts (this probability is over (X, Y) and possibly the provers' randomization).

The *n-times parallel repetition* of the game, is when V chooses $(x_1, y_1), \ldots, (x_n, y_n)$ independently from the distribution (X, Y), gives to P_1 the values $(x_1, \ldots, x_n)$ and to P_2 the values $(y_1, \ldots, y_n)$ to obtain the respective answers $(a_1 \ldots, a_n)$ and $(b_1, \ldots, b_n)$. V accepts only if *all* the answers check out. Note that V acts independently on each instance, but the provers may correlate their answers to different instances.

"Theorem": (Fortnow, Rompel, Sipser 88) If the original game had value at most $(1 - \delta)$ then repeated game has value at most $(1 - \delta)^n$.

"Proof": Obvious— since the all the queries are independent and each one can be satisfied with probability at most $1 - \delta$, the probability that the provers can satisfy all n of them is at most $(1 - \delta)^n$.

Counterexample to "theorem" (Fortnow, Feige): Consider following game: verifier chooses two random and independent bits x, y, the answers are pairs in $\{1, 2\} \times \{0, 1\}$. The verifier accepts the answers a, b iff $a = b = (i, \sigma)$ and Prover i received the bit σ.

Two observations:

1. The value of the game is at most 1/2: for the verifier to accept, the two provers have to send a message with the same i, which means that one prover has to "guess" the bit received by the other prover.
2. The value of the repeated game is at least 1/2: consider the following strategy: first prover on x, x' outputs the answers $(1, x)$ and $(2, x)$ and second prover on input (y, y') will output $(1, y')$ and $(2, y')$. Now if $x = y'$ (which happens with probability 1/2) then the verifier will accept.

Parallel Repetition Theorem (Raz 95, Holenstein 07) The value of the repeated game is at most $2^{-\Omega(\delta^3 n)}$ where the constant in the Ω notation depends (logarithmically) on the alphabet size of the provers' answers.

Proof strategy Let P_1, P_2 be some arbitrary provers. We want to bound the probability of the event $W_1 \wedge W_2 \wedge \cdots \wedge W_n$ (where W_i is the event that they win the i^{th} instance). Since $2^{-\Omega(\delta^3 n)} = (1-\delta)^{\Omega(\delta^2 n)}$, it suffices to show that

$$\text{for all } k \leq c\delta^2 n \;\; \Pr[W_{k+1} | W_1 \wedge \cdots \wedge W_k] \leq 1 - \delta/100 \;\; ,$$

for c some constant depending logarithmically on the provers' alphabet.

In fact, since we can reorder coordinates as we wish, it will suffice to prove the following:

Main Lemma: There's $c > 0$ (depending logarithmically on prover's alphabet) such that for all $k \leq c\delta^2 n$, $\exists j > k$ such that

$$\Pr[W_j | W_1 \wedge \cdots \wedge W_k] \leq 1 - \delta/100 \tag{1}$$

Indeed, since we only want to prove that $W_1 \wedge \cdots \wedge W_n$ is low, it suffices to prove (1) under the assumption that $\Pr[W_1 \wedge \cdots \wedge W_k] \geq 2^{-(k+1)}$, where Σ is the size of the prover's alphabet. This will ensure that letting p_k be the probability of $W_1 \wedge \cdots \wedge W_k$, then $p_{k+1} \leq \max\{2^{-(k+1)}, p_k(1-\delta/100)\}$, which suffices to prove the lemma.

Rough proof idea for Main Lemma: We'll prove the main lemma by reduction. That is, we will assume that there are provers' strategies violating (1) and will use them to succeed in the original (unrepeated) game with probability more than $1 - \delta$.

The idea is that for some coordinate j we will show that if Prover 1 is given x_j and Prover 2 is given y_j with (x_j, y_j) chosen from (X, Y) then Prover 1 is able to sample values $\{x_i\}_{i \neq j}$ and Prover 2 is able to sample values $\{y_i\}_{i \neq j}$ such that the joint distribution of $x_1, \ldots, x_n, y_1, \ldots, y_n$ is statistically close to the distribution of these values conditioned on the first k games succeeding (i.e., conditioned on the event $W_1 \wedge \cdots \wedge W_k$). Therefore, they will be able to win the single game with probability close to $\Pr[W_j | W_1 \wedge \cdots \wedge W_k]$.

This is from now on our focus: how can the two provers perform this sampling.

Easy example: if the provers' messages in one instance do not depend on the questions of another instance then distribution of $\{x_i\}_{i \neq j}$ and $\{y_i\}_{i \neq j}$ is still independent of (x_j, y_j). Hence, the provers can sample from it using shared randomness.

Two useful lemmas: The following two lemmas will be crucial to the proof.

Lemma 1: Let $U = U_1, \ldots, U_n$ be a product distribution and let $\tilde{U} = \tilde{U}_1, \ldots, \tilde{U}_n$ be the distribution of U conditioned on some event that happens with probability at least 2^{-d}. Then $\frac{1}{n}\sum_j \Delta(U_j, \tilde{U}_i) \leq \sqrt{d/n}$.[1]

We'll use Lemma 1 to argue that when $d \ll n$, most indices j satisfy that $\Delta(U_j, \tilde{U}_j)$ is small.

Proof: We prove the lemma for the case that for all i, U_i is the uniform distribution on $\{0,1\}^\ell$ for some ℓ (this is without loss of generality since we can map $\{0,1\}^\ell$ to arbitrary

[1] $\Delta(X, Y)$ denotes the statistical distance of X and Y: $\Delta(X, Y) = 1/2 \sum_z |\Pr[X = z] - \Pr[Y = z]|$.

distribution within $\sim 2^{-\ell}$ accuracy). Let $H()$ denote the Shannon entropy function. Then $H(U) = n\ell$ and $H(\tilde{U}) \geq H(U) - d$. But, since

$$\sum_{j=1}^{n} H(\tilde{U}_j) \geq H(\tilde{U})$$

we get that

$$\tfrac{1}{n}\sum_{j=1}^{n} H(\tilde{U}_j) \geq \ell - d/n$$

Thus if we let $\delta_i = \ell - H(\tilde{U}_j)$ then

$$\tfrac{1}{n}\sum_{j=1}^{n} \delta_i \leq d/n \tag{2}$$

The following fact is left as Exercise 2: If X is a distribution over $\{0,1\}^\ell$ with $H(X) \geq \ell - \delta$ then $\Delta(X, U_\ell)^2 \leq \delta$.

Using it, (2) implies

$$\tfrac{1}{n}\sum_{j=1}^{n} \Delta(\tilde{U}_j, U_j)^2 \leq d/n$$

which using $E[X^2] \geq (E[X])^2$ implies

$$\left(\tfrac{1}{n}\sum_{j=1}^{n} \Delta(\tilde{U}_j, U_j)\right)^2 \leq d/n$$

and taking square roots completes the proof. □

We'll also use a version of Lemma 1 where there might be an additional variable T that is correlated with $U_1, \ldots, U_n$:

Lemma 1': Let $U = U_1, \ldots, U_n$ and T be correlated r.v.'s such that for every $t \in \mathsf{Supp}(T)$, $U|t$ is a product distribution and let $\tilde{U}$ be the distribution of U conditioned on some event W that happens with probability at least 2^{-d}. Then, $\frac{1}{n}\sum_j \Delta(\tilde{T}U_j|\tilde{T}, \tilde{T}\tilde{U}_j) \leq \sqrt{d/n}$, where $\tilde{T} = T|W$.

Explanation of terms in Lemma 1': The distribution $\tilde{T}\tilde{U}_j$ is defined as follows: choose $(t, u_1 \ldots u_n)$ from the correlated random variables T, U conditioning on W, and output tu_j.

The distribution $\tilde{T}U_j|\tilde{T}$ is defined as follows: choose t from T conditioned on W, then choose $u_1 \ldots u_n$ from U conditioned on $T = t$ but *not* on W. Then output tu_j.

Proof of Lemma 1' The proof follows by applying Jensen's inequality to Lemma 1.

$$\tfrac{1}{n}\sum_j \Delta(U_j \circ T, \tilde{U}_j \circ T) = \tfrac{1}{n}\sum_j\sum_t \Pr[T=t|W]\sum_u |\Pr[U_j = u|T=t] - \Pr[\tilde{U}_j = u|T=t| =$$

$$\sum_t \Pr[T=t|W]\tfrac{1}{n}\sum_j \Delta(\tilde{U}_j|T=t, U_j|T=t) \le \text{by Lemma 1}$$

$$\sum_t \Pr[T=t|W]\sqrt{\tfrac{\log(1/\Pr[W|T=t])}{n}} \le$$

$$\tfrac{1}{\sqrt{n}}\sqrt{\log\left(\sum_t \tfrac{\Pr[T=t|W]}{\Pr[W|T=t]}\right)}$$

where the last inequality follows from the fact that $E[f(x)] \le f(E[x])$ for a concave function f, and $x \mapsto \sqrt{\log x}$ is concave for $x > 1$.

But $\frac{\Pr[T=t|W]}{\Pr[W|T=t]} = \frac{\Pr[T=t]}{\Pr[W]}$ and hence this sum is equal to

$$\tfrac{1}{\sqrt{n}}\sqrt{\log\left(\tfrac{1}{\Pr[W]}\sum_t \Pr[T=t]\right)} = \sqrt{\tfrac{\log(1/\Pr[W])}{n}}$$

□

Lemma 2: There is a method for two provers with shared randomness to perform the following: Prover 1 is given a specification of a distribution D and Prover 2 a specification of D', where $\Delta(D, D') \le \epsilon$. Then, letting d and d' be the outputs of Prover 1 and Prover 2 respectively, **(1)** the distribution of d is within ϵ statistical distance to D and **(2)** $\Pr[d = d'] \ge 1 - 2\epsilon$.

Proof: First, it's worthwhile to note that the obvious procedure to this when $D \equiv D'$ (choose a random $p \in [0,1]$ and output the first i such that $\sum_{j \le i} \Pr[D - j] \ge p$) completely breaks down if D' is even slightly different than D.

Consider the case where D and D' are flat distributions. In this case, we can think of them as sets with symmetric difference at most 2ϵ compared to their size. The provers use their shared randomness to take a random ordering of the universe. Prover 1 will output the minimal element in D according to this ordering, and Prover 2 will output the minimal element in D' according to this ordering. The probability that the minimal element falls inside the shared intersection is at least $1 - 2\epsilon$.

This case is actually general, since we can make D and D' flat by considering the distribution (x, p) over all pairs (x, p) such that $x \in \mathsf{Supp}(D)$ and $p \in [0, \Pr[D = x]]$. □

Proof of the Main Lemma: let $k \le \frac{\delta^2 n}{\log|\Sigma| 10^6}$ (where Σ is the alphabet used in the provers' responses) and assume (1) is false for some prover strategies P_1, P_2.

Let's fix $\bar{x} = (x_1, \ldots, x_k), \bar{y} = (y_1, \ldots, y_k), \bar{a} = (a_1, \ldots, a_k), \bar{b} = (b_1, \ldots, b_k)$ to be a "typical" winning transcript of the first k rounds (i.e., a transcript that causes the verifier to accept all these k rounds). By "typical" we mean the following:

- Conditioned on the queries and answers falling in this transcript, 90% of the coordinates j satisfy that W_j with probability at least $1-\delta/2$. Since we assume (1) is false, this will happen with at least 0.9 probability.
- Conditioned on the verifier's first k queries being $\bar{x}, \bar{y}$, the probability that the two provers will answer with $\bar{a}, \bar{b}$ is at least $|\Sigma|^{-4k}$, where Σ is the alphabet used in the provers' responses. (Because the provers' responses in the first k rounds may depend on the verifier's queries in different rounds, this probability is over the choice of $(x_{k+1}, y_{k+1}), \ldots, (x_n, y_n)$.)

  Indeed, because that there are at most $|\Sigma|^{2k}$ possible answers, with 0.9 probability, a random winning transcript will satisfy that the answers $\bar{a}, \bar{b}$ are obtained with probability at least $\Pr[W_1 \wedge \cdots \wedge W_k]/(100|\Sigma|^{2k})$, and we assume that $\Pr[W_1 \wedge \cdots \wedge W_k] \geq |\Sigma|^{-k}$.

Define the distribution $(X_1, Y_1), \ldots, (X_n, Y_n)$ as follows: for $i \leq k$, $X_i = x_i$ and $Y_i = y_i$ with probability 1, and for $i > k$, (X_i, Y_i) is distributed independently according to (X, Y). Note that that is a product distribution (each pair (X_i, Y_i) is independent of the other pairs).

Let W denote the event that the two provers' responses are $\bar{a}$ and $\bar{b}$ respectively, and let $(\tilde{X}_1, \tilde{Y}_1), \ldots, (\tilde{X}_n, \tilde{Y}_n)$ be the distribution of $(X_1, Y_1), \ldots, (X_n, Y_n)$ conditioned on W. Note that this is no longer necessarily a product distribution.

Definition of T: We now make the following crucial definition of a random variable T that is correlated with $(X_1, Y_1), \ldots, (X_n, Y_n)$. Given a sequence of values, $x_1, y_1, \ldots, x_n, y_n$, for every $j > k$, we define the random variable T_j correlated with (x_j, y_j) as follows: with probability 1/2 we set $T_j = (\texttt{'x'}, x_j)$ and with probability 1/2 we set $T_j = (\texttt{'y'}, y_j)$. We define the variable T to be the concatenation of T_j for all $j > k$ and the variable $T_{\setminus j}$ to be the concatenation of T_i for $i \neq j$. We denote $\tilde{T}_j$ to be the distribution of T_j conditioned on W, and define similarly $\tilde{T}$,$\tilde{T}_{\setminus j}$.

The lemma will follow from the following claim:

Claim: Let $\epsilon = \delta/10$. Then for 80% of the coordinates $j > k$:

$$(\tilde{X}_j, \tilde{Y}_j) \approx_\epsilon (X, Y) \tag{3}$$

$$\tilde{T}_{\setminus j} | \tilde{X}_j, \tilde{Y}_j \approx_\epsilon \tilde{T}_{\setminus j} | \tilde{X}_j \tag{4}$$

$$\tilde{T}_{\setminus j} | \tilde{X}_j, \tilde{Y}_j \approx_\epsilon \tilde{T}_{\setminus j} | \tilde{Y}_j \tag{5}$$

The magic of the proof (in both Raz and Holenstein's paper) is in this claim. (4) and (5) deserve some explanation. For example (4) means that if we choose $(\tilde{x}_j, \tilde{y}_j)$ from $(\tilde{X}_j, \tilde{Y}_j)$, then the expected statistical distance of the following two distributions is at most ϵ:

- Choose $\tilde{x}_i, \tilde{y}_i$ for $i \neq j$ according to $\tilde{X}_1, \tilde{Y}_1, \ldots, \tilde{X}_n, \tilde{Y}_n | \tilde{X}_j = \tilde{x}_j, \tilde{Y}_k = \tilde{y}_j$. Choose t from $\tilde{T}_{\setminus j}$ conditioned on $\tilde{x}_1, \tilde{y}_1, \ldots, \tilde{x}_n, \tilde{y}_n$. Output t.
- Choose y_j conditioned on $\tilde{x}_j$ only (i.e. y_j does not necessarily equal $\tilde{y}_j$). Choose $\tilde{x}_i, \tilde{y}_i$ for $i \neq j$ according to $\tilde{X}_1, \tilde{Y}_1, \ldots, \tilde{X}_n, \tilde{Y}_n | \tilde{X}_j = \tilde{x}_j, \tilde{Y}_k = y_j$. Choose t from $\tilde{T}_{\setminus j}$ conditioned on $\tilde{x}_1, \tilde{y}_1, \ldots, \tilde{x}_j, y_j, \ldots, \tilde{x}_n, \tilde{y}_n$. Output t.

Proof of Lemma from Claim Choose j at random and fix it. The two provers will use the following algorithm:

1. Verifier chooses (x,y) from (X,Y). Prover 1 gets x and sets $x_j = x$ and Prover 2 gets y and sets $y_j = y$.
2. Both provers set $x_1, y_1, \ldots, x_k, y_k$ according to $\bar{x}$, $\bar{y}$.
3. Prover 1 computes the distribution $D = \tilde{T}_{\setminus j}|x_j$ and Prover 2 computes the distribution $D' = \tilde{T}_{\setminus j}|y_j$. Due to (4) and 5 we know these two distributions are close to one another and to $\tilde{T}_{\setminus j}|x_j, y_j$. They both use the procedure of Lemma 2 to sample the same value t from this distribution.
4. Recall that t contains for every $i > k$ with $i \neq j$ a pair (d_i, z_i) where $d_i \in \{1,2\}$. Let S be the set of i's such that $d_i = \text{`}\mathtt{x}\text{'}$, and $\bar{S}$ the sets of $i$'s such that $d_i = \text{`}\mathtt{y}\text{'}$. Both provers set $x_i = z_i$ for $i \in S$ and $y_i = z_i$ for $i \in \bar{S}$. At this point for every i, each prover has a value for at least one of x_i or y_i.
5. Prover 1 chooses x_i for all $i \in S$ in the following way: it chooses x_i according to the distribution $X|Y = y_i$, and then it conditions on the first k values of $P_1(x_1, \ldots, x_n)$ being $\bar{a}$ (i.e., if this condition doesn't hold the prover makes all choices again). Similarly for all $i \in \bar{S}$, Prover 2 chooses y_i conditioned on x_i and then conditions on the first k values of $P_2(y_1, \ldots, y_n)$ being $\bar{b}$.
6. Prover 1's answer is the j^{th} value of $P_1(x_1, \ldots, x_n)$. Similarly, Prover 2's answer is the j^{th} value of $P_2(x_1, \ldots, x_n)$.

The key observation is that if (3),(4), and (4) held with $\epsilon = 0$ then the values $x_{k+1}, \ldots, x_n$,$y_{k+1}, \ldots, y_n$ would be exactly distributed according to the distribution $\tilde{X}_{k+1}, \ldots, \tilde{X}_n, \tilde{Y}_{k+1}, \ldots, \tilde{Y}_n$, and so the provers will convince the verifier with probability $1 - \delta/2$. Now, since they hold approximately, they will still convince the verifier with probability at least $1 - \delta/2 - 5\epsilon > 1 - \delta$, leading to a contradiction. □

Proof of Claim We now prove the claim. By Lemma 1 for 90% of the coordinates j the statistical distance of $\tilde{X}_j, \tilde{Y}_j$ from the original distribution (X,Y) is at most $\sqrt{k/n} = \delta/100 = \epsilon/10$ and hence (3) is satisfied.

By Lemma 1', for 90% of the j's, the statistical distance of $\tilde{T}, \tilde{X}_j, \tilde{Y}_j$ from $\tilde{T}, (X_j, Y_j)|\tilde{T}$ is also at most $\epsilon/10$. Since $\tilde{T}$ is equal to $\tilde{T}_{\setminus j}(\text{`}\mathtt{x}\text{'})\tilde{X}_j$ with probability 1/2 and $\tilde{T}_{\setminus j}(\text{`}\mathtt{y}\text{'})\tilde{Y}_j$ with probability 1/2,

$$1/2\Delta(\tilde{T}_{\setminus j}\tilde{X}_j\tilde{Y}_j, \tilde{T}_{\setminus j}\tilde{X}_j\, Y_j|\tilde{X}_j) + 1/2\Delta(\tilde{T}_{\setminus j}\tilde{X}_j\tilde{Y}_j, \tilde{T}_{\setminus j}X_j|\tilde{Y}_j\, \tilde{Y}_j) \leq \epsilon/10$$

implying that

$$\tilde{T}_{\setminus j}\tilde{X}_j\tilde{Y}_j \approx_{\epsilon/5} \tilde{T}_{\setminus j}\tilde{X}_j\, Y_j|\tilde{X}_j \tag{6}$$

$$\tilde{T}_{\setminus j}\tilde{X}_j\tilde{Y}_j \approx_{\epsilon/5} \tilde{T}_{\setminus j}X_j|\tilde{Y}_j\, \tilde{Y}_j \tag{7}$$

Yet (6) and (7) imply (4) and (5) respectively. (6) implies (4) since if (4) was false there would be a distinguisher that on input x, y that are selected from $\tilde{X}_j, \tilde{Y}_j$ and a third input t manages to tell apart if t is selected from $\tilde{T}_{\setminus j}|x$ or t is selected from $\tilde{T}_{\setminus j}|x, y$.[2] But an equivalent way to describe this is that the distinguisher is given x, y, t that either come from $\tilde{X}_j, \tilde{Y}_j, \tilde{T}_{\setminus j}$, or y is only chosen conditioned on x and ignoring t (we use here the fact that by (3) choosing y according to $Y|X = x$ is the same as choosing y according to $\tilde{Y}_j|\tilde{X}_j = x$). Such a distinguisher violates (6). The proof that (7) implies (5) is symmetric. □

[2] We use the fact that we can always flip the output of a distinguisher violating (4) to ensure that it's more likely to output 1 on the left hand side distribution.

Homework Assignments

§1 (25 points)

(a) Given the PCP Theorem as a black-box ($\mathbf{NP} \subseteq \mathbf{PCP}(O(logn), O(1))$) prove that for any language in **NP** there exists a two query PCP proof system where verifier uses $O(\log n)$ randomness and the prover's answers are in an alphabet of constant size, with perfect completeness, and soundness parameter at most ρ for some constant $\rho < 1$. (By soundness parameter we mean the maximum probability of accepting a false statement.)

(b) Using the Parallel Repetition Theorem, show that for every **NP**-language and $\epsilon > 0$, there exists such a system with soundness parameter at most ϵ.

§2 (30 points) Recall that the entropy of a distribution $(p_1, \ldots, p_N)$ is defined to be $\sum_i p_i \log(1/p_i)$. Prove that for every $\delta \in (0, 1)$, if X is a distribution over $\{0, 1\}^\ell$ with $H(X) \geq \ell - \delta$ then $\Delta(X, \mathcal{U})^2 \leq 100\delta$ (this is true even if the constant 100 is replaced by 1, but this version suffices for the proof of the parallel repetition theorem). See footnote for hint[3]

§3 (25 points) Complete the proof of Lemma 1, by showing that the case of general distributions reduces to the case of the uniform distribution.

§4 (30 points) Write down the full proof of the Main Lemma from the claim:

(a) Prove that if (3),(4) and (5) held with $\epsilon = 0$, then the distribution of values $x_1, \ldots, x_n, y_1, \ldots, y_n$ the two provers pick is exactly equal to the distribution $\tilde{X}_1, \ldots, \tilde{X}_n, \tilde{Y}_1, \ldots, \tilde{Y}_n$.

(b) Prove that generally, using (3),(4) and (5) , the distribution of values $x_1, \ldots, x_n, y_1, \ldots, y_n$ the two provers pick is within 10ϵ statistical distance to the distribution $\tilde{X}_1, \ldots, \tilde{X}_n, \tilde{Y}_1, \ldots, \tilde{Y}_n$.

[3] **Hint:** If p_i is the probability that $X = i$, then write $p_i = 2^{-\ell}(1 + x_i)$ and use the estimate $\log(1 + x) \sim x$ for small x's.