A Survey of Parallel Repetition and Projection Games

Abstract

Parallel Repetition governs the decay of the verification probability in a repeated 2-prover interactive protocol. We will first analyze Feige's game which shows the naive exponential bound does not hold. Then we will state Raz's original bound[6] and observe some interesting properties of it. Though the original proof by Raz is very involved, we will give a high level proof idea and analyze major steps involved in Holenstein's simplification as discussed by Rao[4]. Afterwards we will look at a variant of the bound recently discovered by Dinur and Steurer[2] when applied to projection games and compare it with that discovered by Rao. Critical definitions and lemmas are analyzed so that the reader can attain an intuitive reasoning for the derivation of these bounds. This is especially important in the case of Dinur and Steurer since they take a spectral graph theory approach but still utilize the notion of correlated sampling observed in Holenstein's simplification. Afterwards we discuss applications and tightness of some parallel repetition bounds.

1 Introduction

Suppose Alice flips a fair coin n times and she wins a dollar if and only if she guesses every outcome correctly. Clearly the expected value of this game is $\frac{1}{2^n}$. Now suppose Bob also flips n fair coins each but he can't share the results with Alice. Both of them submit "answers" x_i and y_i respectively and in the range $\{0, 1\}$. If the flip results were r_i and q_i, they win on flip i if and only if $r_i + x_i = q_i + y_i \bmod 2$. It is not hard to see in this case either that no matter what strategy they decide to use, there is a $\frac{1}{2}$ probability of winning on each flip and $\frac{1}{2^n}$ probability of winning all flips. In general if some game $\mathcal{G}$ had some winning probability via strategy σ, which we denote as the value of the game or $\text{val}(\mathcal{G})$, then $\mathcal{G}^n$, the game repeated n times, has value $\text{val}(\mathcal{G}^n) \geq \text{val}^n(\mathcal{G})$ since we can apply σ on each game. The next question to ask is if this bound tight i.e. $\text{val}(\mathcal{G}^n) = \text{val}^n(\mathcal{G})$. This is true for the simple two player game we mentioned earlier but it is not true in general for games of the following form.

Definition 1.1. A two player game $\mathcal{G}$ with $\text{val}(\mathcal{G})$ and n-fold repeated game $\mathcal{G}^n$ with $\text{val}(\mathcal{G}^n)$ refers to the following process.

- *A referee asks Player A and Player B questions X and Y respectively where (X, Y) is chosen from some distribution.*
- *Players A and B respond with answers α and β respectively.*
- *The referee evaluates some predicate $V(X, Y, \alpha, \beta)$ and the game is won if the predicate is satisfied.*
- *$\mathcal{G}$ is a single iteration of the above and $\text{val}(\mathcal{G})$ is the maximum winning probability over answer strategies (σ, π). If the strategies are randomized, we take the expected value.*

- *$\mathcal{G}^n$ involves n simultaneous plays of the game where the referee asks questions (X_i, Y_i) each independently drawn from the same distribution, the players respond with answers α_i and β_i. The referee evaluates the predicate V for all $(X_i, Y_i, \alpha_i, \beta_i)$ and* $\text{val}(\mathcal{G}^n)$ *is the maximum probability the predicate is satisfied for all i over all strategies $((\sigma_1, \pi_1), \ldots, (\sigma_n, \pi_n))$. If the strategies are randomized, we take the expected value.*

Note that the game can allow for a publicly shared random string as the value can be matched by a deterministic strategy. We will use A and B interchangeably with Alice and Bob from now on. If the two players can develop a strategy such that winning earlier games can influence later games, then the value can actually exceed the trivial bound. We indeed observe this in Feige's game.

Proposition 1.2. (Feige's Game) The following two player game $\mathcal{G}$ satisfies $\text{val}(\mathcal{G}) = \frac{1}{2}$ and $\text{val}(\mathcal{G}^2) = \frac{1}{2}$.

- *The referee sends A and B random bits X and Y.*
- *The players respond with answers of the form $\alpha, \beta \in \{1, 2\} \times \{0, 1\}$ and win if and only if $\alpha = \beta = (1, X)$ or $\alpha = \beta = (2, Y)$.*

Proof. For $\mathcal{G}$ we can have Alice respond with $(1, X)$ while Bob responds with $(1, Y)$. We have $X = Y$ with probability $\frac{1}{2}$ so the value of the game is $\frac{1}{2}$. For $\mathcal{G}^2$ we let

$$\alpha_1 = (1, X_1) \qquad \alpha_2 = (2, X_1)$$
$$\beta_1 = (1, Y_2) \qquad \beta_2 = (2, Y_2)$$

Since $X_1 = Y_2$ with probability $\frac{1}{2}$, $\text{val}(\mathcal{G}^2) = \frac{1}{2}$. □

Feige's Game is a simple result regarding parallel repetition; it is not immediate obvious whether there is exponential decay in the game's value with respect to n. However if one were to repeat the above game for $n \geq 3$ we would eventually observe the expected exponential decay. Raz proved the law governing the decay which was then simplified by Holenstein to give the following result for general games:

Theorem 1.3. (Parallel Repetition Theorem) [Raz98, Hol07] *If* $\text{val}(\mathcal{G})$ *is at most $1 - \epsilon$ then there is a universal constant $\gamma > 0$ such that the value of $\mathcal{G}^n$ is at most $(1 - \epsilon)^{\gamma \epsilon^2 n / c}$ where c is a bound on the answer length.*

Raz's original proof was very complicated and was one of the first applications of information theory in complexity theory. We will give an overview of the derivation of Holenstein's simplification in this survey. These results and other related bounds utilize the observation that a strategy which results in a high probability of winning on a small subset of the repeated games implies the strategy does worse on some game not in that subset. Feige's counterexample circumvents this problem for $n = 2$ but the parallel repetition theorem says it will fall victim to exponential decay although at a non-obvious rate.

The parallel repetition theorem has very deep connections to complexity theory so it has been of recent importance. The game is essentially a protocol involving two provers and a verifier and we can use parallel repetition improve soundness via repetition in certain proof systems[3]. We can also reformulate these games as graph representations of Constrained Satisfaction Problems (CSPs).

Definition 1.4. A two player game $\mathcal{G}$ with $\text{val}(\mathcal{G})$ can be represented via a bipartite graph with edge set $E \subseteq V_1 \times V_2$. We associate each edge (X, Y) with a constraint set $\mathcal{V}_{X,Y}$ and redefine the game as follows.

- *A referee asks Player A and Player B sends X and Y respectively where (X, Y) is an edge chosen from some distribution.*
- *Players A and B respond with answers α and β respectively.*
- *The referee checks if $(\alpha, \beta) \in \mathcal{V}_{X,Y}$ and the game is won if so.*
- *$\mathcal{G}$ is a single iteration of the above and $\text{val}(\mathcal{G})$ is the maximum fraction of edge constraints satisfied over vertex assignments σ and π on V_1 and V_2 respectively. If the assignments are randomized, we take the expected value.*
- *$\mathcal{G}^n$ involves n simultaneous plays of the game where the referee chooses edges (X_i, Y_i) independently and uniformly, the players respond with answers α_i and β_i. The referee checks if $(\alpha_i, \beta_i) \in \mathcal{V}_{X_i,Y_i}$ and game i is won if so. $\mathcal{G}^n$ is won if all the games are won and as a result $\text{val}(\mathcal{G}^n)$ is the fraction of constraints satisfied for each game over assignments $((\sigma_1, \pi_1), \ldots, (\sigma_n, \pi_n))$ where a constraint in this sense is of the form $\mathcal{V}^{\otimes n}$. If the assignments are randomized, we take the expected value.*

The value of this CSP is equivalent to $\text{val}(\mathcal{G}^n)$. We know from the PCP Theorem, that there exists some constant α such that obtaining an α-approximation to some CSP instance is NP-hard. We can actually strengthen this using Parallel Repetition.

Theorem 1.5. (PCP + Parallel Repetition) For every $\epsilon > 0$, there exists a CSP that is NP-hard to ϵ-approximate.

Proof. Amplify the α-approximation CSP and by considering the value of its product graph G^n and apply the parallel repetition theorem. □

The CSP view of Parallel Repetition is intuitive and Dinur and Steurer use it to introduce the notion of symmetrized games which they use in their derivation of a parallel repetition bound on *projection games.* A projection game is a game where for each constraint on an edge, there is most one acceptable answer from Alice given Bob's answer.

Theorem 1.6. (Projection Games + Parallel Repetition) [DS14] If the value of a projection game $\mathcal{G}$ is upper bounded by ρ then

$$\text{val}(\mathcal{G}^n) \leq \left(\frac{2\sqrt{\rho}}{1+\rho}\right)^{n/2}$$

Rao also derived a bound for projection games using the information theoretic approach but Dinur and Steurer's bound uses spectral graph theory and has stronger applications making it interesting to study.

2 Parallel Repetition

2.1 Statistical Distributions

The proof of the parallel repetition theorem focuses on analyzing the transformations of the distributions of successes as we learn more about the results of the repeated games. We will need to introduce some theory regarding certain types of distributions and their relationships to each other. The distribution of a random variable follows the standard notion of a probability distribution, containing information on the random variable's probability of taking some instance. The *support* of a distribution will be denoted by supp and it is the set of elements which map to a positive value under the probability measure.

Many types of distribution will pop up in our analysis so we will need the following notation to make things convenient. Assume we are given distributions A, B,and C and for each of the following product distributions we provide the rule for its measure given the measures for events involving A, B, and C.

- $A \otimes B \otimes C$
$$\mu(A \otimes B \otimes C) = \mu_A(A)\mu_B(B)\mu_C(C)$$
- $AB \otimes C$
$$\mu(AB \otimes C) = \mu_{A,B}(A, B)\mu_C(C)$$
- $AB \otimes (C|B)$
$$\mu(AB \otimes (C|B)) = \mu_{A,B}(A, B)\mu_{C|B}(C|B)$$

We will now state some facts regarding the *closeness* of distributions.

Definition 2.1. Let D and F be two random variables taking values from a common set $\mathcal{S}$. Their *statistical distance* $\Delta(D, F)$ is defined as

$$\begin{aligned}\Delta(D, F) &= \max_{T \subseteq \mathcal{S}}(|\Pr[D \in \mathcal{T}] - \Pr[F \in \mathcal{T}]|) \\ &= \frac{1}{2}\sum_{s \in \mathcal{S}} |\Pr[D = s] - \Pr[F = s]|\end{aligned}$$

Furthermore we say D is *ϵ-close* to F if $\Delta(D, F) \leq \epsilon$ which we will denote by $D \overset{\epsilon}{\approx} F$. The following propositions are intuitive and follow easily.

Proposition 2.2. If $D \overset{\epsilon}{\approx} F$ and g is any function on $\mathcal{S}$, then $g(D) \overset{\epsilon}{\approx} g(F)$.

Proposition 2.3. If $D \overset{\epsilon_1}{\approx} F$ and $F \overset{\epsilon_2}{\approx} G$ then $D \overset{\epsilon_1+\epsilon_2}{\approx} G$.

Proposition 2.4. Suppose E_1 and E_2 are events such that $\Pr[E_1] = \Pr[E_2] = \lambda$, then $\Delta(D|E_1, F|E_2) \leq \Delta(D, F)/\lambda$.

Proposition 2.5. If $\Pr[D \neq F] \leq \epsilon$, then $D \overset{\epsilon}{\approx} F$.

We observe that the closer two distributions are according to statistical distance, the more likely they are to match. This will be our main tool of observing how distributions are related but we will need to briefly analyze a related quantity.

Definition 2.6. The *informational divergence* between D and F is defined as

$$\nabla(D, F) = \sum_{s \in \mathcal{S}} \Pr[D = s] \log \left(\frac{\Pr[D = s]}{\Pr[F = s]} \right)$$

The above definition assumes $0 \log 0 = 0$ and that if $\frac{\Pr[D=s]}{\Pr[F=s]}$ is not defined then $\nabla(D, F) = \infty$. Note that the informational divergence is not symmetric, unlike the statistical distance, and models the expected information lost when the distribution F is used in place of D. We easily obtain the following facts

Fact 2.7. $\nabla(D, F) \geq [\Delta(D, F)]^2$

Fact 2.8. If E is an event such that $\Pr[E] = 2^{-d}$, then $\nabla(D|E, D) \leq d$.

Proof.

$$\begin{aligned}
\nabla(D|E, D) &= \sum_{s \in \mathcal{S}} \Pr[D = s|E] \log \left(\frac{\Pr[D = s|E]}{\Pr[D = s]} \right) \\
&= \sum_{s \in \mathcal{S}} \Pr[D = s|E] \log \left(\frac{\Pr[D = s, E]}{\Pr[D = s] \Pr[E]} \right) \\
&= \sum_{s \in \mathcal{S}} \Pr[D = s|E] \log(1/\Pr[E]) + \sum_{s \in \mathcal{S}} \Pr[D = s|E] \log \left(\frac{\Pr[D = s, E]}{\Pr[D = s]} \right) \\
&\leq \log(1/\Pr[E]) \qquad \text{follows from } \Pr[D = s, E] \leq \Pr[D = s]
\end{aligned}$$

□

Fact 2.9. If $D_1, D_2, \ldots, D_n$ are independent random variables and $F_1, F_2, \ldots, F_n$ are some other random variables, then

$$\sum_{i=1}^{n} \nabla(F_i, D_i) \leq \nabla(F_1 \ldots F_n, D_1 \ldots D_n)$$

where $X_1 \ldots X_n$ denotes the distribution of $(X_1, \ldots, X_n)$.

Proof. Since the D_i's are independent, the amount of information we lose when we consider the combined distribution of D_i's over the F_i's is greater than the sum of the individual losses. □

We are now ready to prove a core lemma in Raz's initial paper.

Lemma 2.10. Let $D_1, D_2, \ldots, D_n$ be independent random variables and E be an event such that $\Pr[E] = 2^{-d}$, then

$$\mathop{\mathbb{E}}_{i}[\Delta(D_i|E, D_i)] \leq \sqrt{\frac{d}{n}}$$

Proof.

$$\begin{aligned}
\mathop{\mathbb{E}}_{i}[\Delta(D_i|E, D_i)]^2 &\leq \mathop{\mathbb{E}}_{i}[\Delta^2(D_i|E, D_i)] \\
&\leq \mathop{\mathbb{E}}_{i}[\nabla(D_i|E, D_i)] \\
&= \frac{1}{n}\sum_i \nabla(D_i|E, D_i) \\
&\leq \frac{1}{n}\nabla(D_1 \ldots D_n|E, D_1 \ldots D_n) \\
&\leq \frac{d}{n}
\end{aligned}$$

□

We will find this useful in the proof of Parallel Repetition as if we interpret the D_i as a distribution representing game i, the lemma implies the existence of some game i whose corresponding distribution doesn't get altered significantly when conditioned on a dense event e.g. some game having a high winning probability. The following corollary also follows from a bit of menial algebra and applying Lemma 2.10

Corollary 2.11. Let $A, R, D_1, \ldots D_n$ be random variables and E be an event with $\Pr[E] = 2^{-d}$ such that the following hold:

- For every r, $D_1, \ldots, D_n$ are independent conditioned on $R = r$.
- For every r, $|\operatorname{supp}(A|E \wedge (R = r))| \leq 2^h$.

Then,

$$\mathop{\mathbb{E}}_{i}[\Delta\big((RA|E) \otimes (D_i|R), RAD_i|E\big)] \leq \sqrt{\frac{d+h}{n}}$$

Now assuming that d is small i.e. E is dense and h is small i.e. A has small support given E and R, we can say for some i

$$\Pr(D_i|R) \approx \frac{\Pr(RAD_i|E)}{\Pr(RA|E)} = \frac{\Pr(RAD_iE)}{\Pr(RAE)} = \Pr(D_i|RAE)$$

so the earlier claim that some distribution D_i barely changes can be strengthened to only rely on conditional independence on some random variable and we can further condition on a random variable with small support.

2.2 Proof of Parallel Repetition

As we mentioned before, we want to first bound the probability of winning the remaining games given successes on a set of games $S \subset [n]$ in $\mathcal{G}^n$. We shall denote W_S to be the event all games in S are won and appeal to the following lemma

Lemma 2.12. Let S be of size k. If $\mathcal{G}$ is such that $\operatorname{val}(\mathcal{G}) = 1 - \epsilon$, the answers are of length c, and $\mathbb{P}[W_S] \geq 2^{-\frac{\epsilon^2(n-k)}{34^2} + kc}$, then $\mathbb{E}_{i \notin S}[\mathbb{P}[W_i|W_S]] \leq 1 - \epsilon/2$.

Before proving this, we need to introduce some random variables regarding the outcomes of the games. Let $S = \{n-k+1, n-k, \ldots, n\}$ without loss of generality be the won indices. Let the answers to these won games be $A = A_{n-k+1} \ldots A_n$ and $B = B_{n-k+1} \ldots B_n$ respectively. We also define the following random variables

- T_i represents a uniformly random question for each $i = 1, 2, 3 \ldots, n-k$ and U_i the opposite question i.e. (T_i, U_i) is chosen from $\{(X_i, Y_i), (Y_i, X_i)\}$ with uniform probability.
- $Q = X_{n-k+1}Y_{n-k+1}X_{n-k+2}Y_{n-k+2} \ldots X_nY_n$ represents the won questions.
- Let $R = QT_1T_2 \ldots T_{n-k}$ represent the won questions and a random question from each of the remaining question pairs.
- Let $R^{-j} = QT_1T_2 \ldots T_{j-1}T_{j+1} \ldots T_n$ represent R but with the j'th coordinate removed.

The previous lemma then can be reduced to the following

Lemma 2.13. Suppose h is a positive integer such that

- $\mathbb{P}[W_S] \geq 2^{-\frac{\epsilon^2(n-k)}{34^2}+h}$
- For every r, $|\operatorname{supp}(A|R = r \wedge W_S)| \leq 2^h$

Then $\mathbb{E}_{i \notin S}[\mathbb{P}[W_i|W_S]] \leq 1 - \epsilon/2$.

Proof of Lemma 2.12. If $\mathcal{G}$ is such that the answers are of length c we have that the support of the distribution of the answers for a particular questions is bounded by 2^c. Since A consists of k questions we have that its support is bounded by 2^{kc}. The lemma then follows by applying $h = kc$ in Lemma 2.13. □

Proof of Lemma 2.13. Before deferring to the proof, let us outline the strategy that will be used. We first observe how to go from a strategy for $\mathcal{G}^n$ to a strategy for $\mathcal{G}$ given W_S. Given questions X and Y they can select an index i such that $(X_i, Y_i) = (X, Y)$ and they generate (X_k, Y_k) for $k \neq i$ via shared randomness. They can then run the strategy for $\mathcal{G}^n$ to win index i. In order for this to be meaningful at all, we would like their sampling to be statistically close to $X^nY^n|W_S$ where all the questions are generated randomly. If the distance to this distribution was $\epsilon/2$, this immediately implies the lemma since otherwise then they can win the i'th coordinate with probability greater than $1 - \epsilon/2 - \epsilon/2 = 1 - \epsilon$. We will observe what happens for the average $i \notin S$. Alice and Bob will sample first sample the two random variables A, R^{-i} (Alice's answers to the won questions along with the won questions and a question from each of the remaining pairs except i). This is useful since it tells us that $(X^n, Y^n|(X_i, R^{-i}, A) \wedge W_S)$ and $(X^n, Y^n|(Y_i, R^{-i}, A) \wedge W_S)$ are both product distributions i.e. X^n and Y^n are independent given those conditions since they fix at least one of (X_k, Y_k) for each k. As a result, they can generate the remaining questions independently and obtain a distribution close to $X^nY^n|W_S$. One critical step to note is that Alice and Bob must obtain equal samples AR^{-i} with high probability and we will need to refer to the following claim to do so.

Claim 2.14. There exists a protocol for l non-communicating players such that given distributions $A_1, \ldots, A_l$ taking values in $\mathcal{A}$ such that $\Delta(A_l, A_i) \leq \epsilon_i$ for all $i \in [l]$, the players can use shared randomness to sample $B_1, \ldots, B_l$ with the following properties:

- *For every i, B_i has the same distribution as A_i.*
- *The probability that the samples are inconsistent is bounded above by $2\sum_{i=1}^{l-1} \epsilon_i$.*

We will see an example of such a protocol in the section on projection games. We will use this to allow the players to obtain samples $AR^{-i}|W_S$ and show they are statistically close to the desired samples. This idea to use shared randomness to generate the questions is referred to as *correlated sampling*. For convenience all the following references to $D_i \overset{\widehat{\epsilon}}{\approx} F_i$ will now refer to the expected value of the distance over $i \notin S$ i.e. $\mathbb{E}_{i \notin S}\Delta(D_i, F_i) \leq \epsilon$. Now we will develop some claims involving the proximity of the given distributions to the desired ones.

Claim 2.15. $X_iY_i|W_S \overset{\widehat{\epsilon/34}}{\approx} X_iY_i$

Proof. Direct application of Lemma 2.10. □

Claim 2.16. $AR^{-i}Y_iX_i|W_S \overset{\widehat{\epsilon/17}}{\approx} (AR^{-i}Y_i|W_S) \otimes (X_i|Y_i)$

Proof. Apply Corollary 2.11 to get $ARU_i|W_S \overset{\widehat{\epsilon/34}}{\approx} (AR|W_S) \otimes (U_i|R) = (AR|W_S) \otimes (U_i|T_i)$. Conditioning on $(T_i, U_i) = (Y_i, X_i)$ and applying Proposition 2.4 we arrive at the claim. □

Claim 2.17. $(X_iY_i) \otimes (AR^{-i}|X_iW_S) \overset{\widehat{4\epsilon/34}}{\approx} (X_iY_i) \otimes (AR^{-i}|X_iY_iW_S) \overset{\widehat{4\epsilon/34}}{\approx} (X_iY_i) \otimes (AR^{-i}|Y_iW_S)$

Proof. We repeatedly apply Claims 2.15 and 2.16 and make necessary rearrangements.

$$\begin{aligned}
(X_iY_i) \otimes (AR^{-i}|Y_iX_iW_S) &\overset{\widehat{\epsilon/34}}{\approx} (X_iY_i|W_S) \otimes (AR^{-i}|Y_iX_iW_S) \\
&= AR^{-i}X_iY_i|W_S \\
&\overset{\widehat{\epsilon/17}}{\approx} (AR^{-i}Y_i|W_S) \otimes (X_i|Y_i) \\
&= (X_i|Y_i) \otimes (Y_i|W_S) \otimes (AR^{-i}|Y_iW_S) \\
&\overset{\widehat{\epsilon/34}}{\approx} (X_i|Y_i) \otimes (Y_i) \otimes (AR^{-i}|Y_iW_S) \\
&= (X_iY_i) \otimes (AR^{-i}|Y_iW_S)
\end{aligned}$$

The other equality follows from a symmetric argument except conditioning on $(T_i, U_i) = (X_i, Y_i)$. □

By running the protocol from Claim 2.14, we have that in expectation over $i \notin S$ the probability the three random variables of Claim 6.3 are inconsistent is bounded by $2(4\epsilon/34) + 2(4\epsilon/34) = 16\epsilon/34$. This implies that the distribution of AR^{-i} for each player is $16\epsilon/34$ close to the distribution of AR^{-i}

conditioning on their joint distribution. We now just need one more application of Claim 2.15 to get rid of this dependence

$$\begin{aligned}&\mathbb{E}_{i\notin S}\Delta((X_iY_i)\otimes(AR^{-i}|X_iY_iW_S), X_iY_iAR^{-i}|W_S)\\&=\mathbb{E}_{i\notin S}\Delta((X_iY_i)\otimes(AR^{-i}|X_iY_iW_S),(X_iY_i|W_S)\otimes(AR^{-i}|X_iY_iW_S))\\&\le\mathbb{E}_{i\notin S}\Delta(X_iY_i, X_iY_i|W_S)\\&\le\epsilon/34\end{aligned}$$

Therefore on average, Alice and Bob are sampling from a distribution that is $17\epsilon/34=\epsilon/2$ close to the correct one, where our random variables are only conditioned on W_S. As we mentioned before, they can generate questions independently but now we know the distribution of the questions they get is statistically close to $X^nY^n|W_S$. In particular we get the following result from the bound on $\mathcal{G}$'s value

$$1-\epsilon\ge\mathbb{E}_{i\notin S}[\Pr[W_i|W_S]]-\epsilon/2$$

which immediately implies the lemma. □

We now conclude with the proof of the Parallel Repetition Theorem.

Proof of Theorem 1.3. We introduce explicit constants, showing that the value of the repeated game is bounded by $(1-\epsilon/2)^t$ where $t=\frac{\epsilon^2 n}{35^2+34^2c}$ which satisfies $\epsilon t\le\frac{\epsilon^2(n-t)}{34^2}-tc$. Suppose for the sake of contradiction that $\mathrm{val}(\mathcal{G}^n)>(1-\epsilon/2)^t$ and that k is the smallest number such that every $H\subset[n]$ of size $k+1$ satisfies $\Pr[W_H]>(1-\epsilon/2)^{k+1}$. This implies the existence of a set $S\subset[n]$ of size k such that $\Pr[W_S]\le(1-\epsilon/2)^k$. We thus have for $\epsilon\in[0,1]$ and $k<t$

$$\Pr[W_S]\ge(1-\epsilon/2)^t\ge 2^{-\epsilon t}\ge 2^{-\frac{\epsilon^2(n-t)}{34^2}+tc}\ge 2^{-\frac{\epsilon^2(n-t)}{34^2}+kc}$$

Applying Lemma 2.12 shows that there exists some i such that $\Pr[W_i|W_S]\le(1-\epsilon/2)$ but then $\Pr[W_{S\cup\{i\}}]\le(1-\epsilon/2)^{k+1}$, contradicting our assumption about H. □

3 Projection Games

3.1 Projection Operators

In the analysis of parallel repetition on projection games, we will assume the bipartite graph interpretation of a game $G=(V_1,V_2,E,\mathcal{V})$. It may also help to extend the graphs so that the vertex includes an answer parameter i.e. Alice's set is (X,α) and Bob's is (Y,β) such that (X,α) is connected to (Y,β) iff $X\in V_1, Y\in V_2$, $(X,Y)\in E$ and $(\alpha,\beta)\in\mathcal{V}_{X,Y}$. The definition of a projection game then says that for each constraint $\mathcal{V}_{X,Y}$ and β there is at most one α such that $(\alpha,\beta)\in\mathcal{V}_{X,Y}$. Alice can assign each vertex a value $g(u,\alpha)$ such that $\sum_\alpha g(u,\alpha)=1$[1], so that $g(u,\alpha)$ represents the probability of producing answer α given u. We define $f(v,\beta)$ similarly for Bob.

The space of g and f each can be associated with an inner product

$$\langle g,g'\rangle=\mathop{\mathbb{E}}_u\sum_\alpha g(u,\alpha)g'(u,\alpha)\qquad\langle f,f'\rangle=\mathop{\mathbb{E}}_v\sum_\beta f(v,\beta)f'(v,\beta)$$

[1] This isn't necessary but it conveniently gets rid of normalization factors

These can be interpreted as a measure of the overlap of two assignments averaged over the vertices. We also use the standard definition of norm, $\|f\| = \langle f, f\rangle^{1/2}$. For a projection game, we will use the notation $\beta_{\mathcal{V}} \sim \alpha$ to represent the set of β incident to α for constraint $\mathcal{V}$ while $v \sim u$ represents the v incident to u in E. The game can then be identified by the following linear operator where $\mathcal{V} = \mathcal{V}_{u,v}$

$$Gf(u, \alpha) = \mathop{\mathbb{E}}_{v \sim u} \sum_{\beta_{\mathcal{V}} \sim \alpha} f(v, \beta)$$

This is essentially finding the expected value over the Alice's vertices of Bob's assignments which project to it. A closer look tells us that the value of a strategy is given by

$$\langle g, Gf\rangle = \mathop{\mathbb{E}}_{(u,v)} \sum_{\alpha, \beta_V \sim \alpha} g(u, \alpha) f(v, \beta)$$

So finding the value of a projection game is equivalent to maximizing $\langle g, Gf\rangle$ over f and g. We can extend the operator to the simultaneous play of two projection games G and H over V and V' respectively, denoted by $G \otimes H$.

$$(G \otimes H)f(u, u', \alpha, \alpha') = \mathop{\mathbb{E}}_{\substack{v \sim u \\ v' \sim u'}} \sum_{\substack{\beta_{\mathcal{V}} \sim \alpha \\ \beta'_{\mathcal{V}'} \sim \alpha'}} f(v, v', \beta, \beta')$$

3.2 Collision Value and Symmetrized Games

We now observe a related quantity, $\|Gf\|^2$ the square of the *collision value* of an assignment f. From our definition of the norm, we see that $\|Gf\|^2$ represents the probability over u that two randomly selected labels from randomly selected vertices v, v' incident to u project to some common α i.e. collide. We define the *collision value* of a game G to be $\|G\| = \max_f \|Gf\|$.

Claim 3.1. $\mathrm{val}(G) \le \|G\| \le \mathrm{val}(G)^{1/2}$

Proof. A simple application of Cauchy-Schwarz assuming f and g attain $\mathrm{val}(G)$ shows $\mathrm{val}(G) = \langle g, Gf\rangle \le \|g\| \cdot \|Gf\| \le \|Gf\| \le \|G\|$. Let f now be such that $\|G\|^2$ then the other inequality follows from $\|G\|^2 = \langle Gf, Gf\rangle \le \max_g \langle g, Gf\rangle \le \mathrm{val}(G)$. □

We shall now show that there is a game whose value is equal to $\|Gf\|^2$ and the vertex sets are both equal to V in what we call the *symmetrized constraint graph* G_{sym}. $(v, v') \in V \times V$ are connected by an edge if they are both incident to some vertex u under constraints $\mathcal{V}$ and $\mathcal{V}'$ respectively. If originally the edges are chosen according to distribution μ, we define the distribution μ_{sym} for the edges in G_{sym} by $\mu_{sym}(v, v') = \sum_u \mu(u, v)\mu(u, v')$. The constraint on (v, v') is the set of pairs (β, β') such that there exists some α such that $\beta_V, \beta'_{V'} \sim \alpha$. Thus we win in this game whenever the β answer pair collide to a projected α of some incident vertex u in the original graph. This form is convenient as eliminates Alice from the analysis. It also tells us that if we give each vertex $v \in V$ in G_{sym} an assignment β with probability $f(v, \beta)$ then the value of this assignment (collision probability) is $\|Gf\|^2$. This view also implies the following since the collision value is also a game value.

Claim 3.2. $\|G \otimes H\| \le \|G\|$

3.3 Trivial Games

We briefly define the following basic projection operators which will have interesting behavior when we consider them in repetition with other games. We define $Tf(v, \alpha)$ to be equal $\sum_{\beta} f(v, \beta)$ when $\alpha = 1$ and 0 otherwise. We also define $T_v f(v', \alpha) = Tf(v, \alpha)$ for all $v' \in V$ which essentially restricts Bob's vertex set to $\{v\}$. This gives us a lot of freedom in the assignment for f without altering T_v's value so it has a stationary role when played parallel to another projection game.

3.4 Proof Overview

Dinur and Steurer's bound, Theorem 1.6, reduces to the following

Theorem 3.3. Any two projection games G and H satisfy $\|G \otimes H\|^2 \leq \varphi(\|G\|^2) \cdot \|H\|^2$, where $\varphi(x) = \frac{2\sqrt{x}}{1+x}$,

The reduction is easy to see by Claim 3.1 and applying the above repeatedly. We will now observe the nature of such a multiplicative parameter i.e. ρ_G such that for every projection game H

$$\|G \otimes H\| \leq \rho_G \cdot \|H\|$$

It is easy to note that the minimum value of ρ_G is

$$\rho_G = \sup_H \frac{\|G \otimes H\|}{\|H\|}$$

which is essentially the negative value factor associated with playing G in parallel with any projection game. The naive repetition tells us $\rho_G \geq \|G\|$ but we are interested in whether $\rho_G \approx \|G\|$. This is shown via a relaxed value parameter $\mathrm{val}_+(G) \geq \rho_G$ such that $\mathrm{val}_+(G) \approx \|G\|$. We let f be an optimal assignment for $G \otimes H$ and "factor" $(G \otimes H)f = (G \otimes I)(I \otimes H)f = (G \otimes I)h$ and $(T_v \otimes H)f = (T_v \otimes I)(I \otimes H)f = (T_v \otimes I)h$ where I is the identity operator on the appropriate space We define the relaxed value as follows.

$$\mathrm{val}_+(G) = \max_{h \geq 0} \frac{\|(G \otimes I_\Omega)h\|}{\max_v \|(T_v \otimes I_\Omega)h\|}$$

We can assume that h operates on $V \times \Sigma \times \Omega$, where Σ is the answer set, for some space Ω since G operates on the first component but H is undetermined so its measure space is arbitrary and hence the subscripts under the identity operator. It follows from Claim 3.2 that $\mathrm{val}_+(G) \geq \rho_G$. $\mathrm{val}_+(G)$ actually turns out to be multiplicative as opposed to $\mathrm{val}(G)$ and this together with Claim 3.1 reduces Theorem 3.3 to the contrapositive of the following.

Theorem 3.4. Let G be a game with $\mathrm{val}_+(G)^2 > \rho$, then

$$\mathrm{val}(G) \geq \|G\|^2 \geq \frac{1 - \sqrt{1 - \rho^2}}{1 + \sqrt{1 - \rho^2}}$$

We assume the existence of a non-negative function $f : (V \times \Sigma \times \Omega) \to \mathbb{R}$ such that $\|(G \otimes I)f\|^2 > \rho \max_v \|(T_v \otimes I)f\|^2$. Furthermore the inequality lets us rescale so that $g \leq 1$ and $\max_v \|(T_v \otimes I)g\| = 1$ where the latter is justified since adjusting the measure equally affects both sides. The proof is broken into three parts which simplify the existing assignment, increasing the contribution from G while T_v's remains the same (or decreases).

1. Convert f into a deterministic assignment, where for each ω, v, $f_{v,\omega}(\beta) = f(v, \beta, \omega)$ assigns at most one β a positive value. This is easily done through an expectation argument on a sampling of deterministic assignments to simulate the randomized assignment and the result is an increase in the value of $\mathrm{val}_+(G)^2$.

2. Convert f into a partial assignment, where the values are either 0 or 1. The result is an assignment and a new space Ω' such that $\|(G \otimes I_{\Omega'})f\|^2 > \psi(\rho) = 1 - (1 - \rho^2)^{1/2}$. This uses randomized rounding and a type of Cheeger's inequality from spectral graph theory applied on projection games. This serves to find a set of partial assignments f_ω so that in the Symmetrized Games, we are guaranteed many collisions and hence increasing the value of the game, proportional to $\psi(\rho)$. Not surprisingly, the analysis becomes much more simpler when we consider expanders over general graphs. In the case of G_{sym} being an expander, f ends up only needing to take a single assignment per $\omega \in \Omega$ i.e. it is essentially of the form $f(v, \beta)$ where for each v there is a unique β assignment. Intuitively this works since expanders are highly connected and sparse, so a single assignment already allows for a nice number of collisions. However this is not true in the general case since G could be made from multiple disjoint expander-like graphs and so we may need to make multiple assignments.

3. The last step is to combine the partial results f_ω into an assignment for G with value $\frac{1-\gamma}{1+\gamma}$ where $1 - \gamma = \psi(\rho)$. The idea is related to the notion of correlated sampling, which we recall uses shared randomness to independently instantiate random variables. In the Holenstein's simplification, Alice and Bob use it to guarantee a statistically close to optimal question distribution so that they can run the G^n strategy to obtain a good strategy for G, exploiting the fact that $\Pr[W_i|S]$ is decent for some i and that sampling AR^{-i} gets us a product distribution. In this context, we started out with a lower bound on the relaxed value $\mathrm{val}_+(G)^2 > \rho$ and are now using it to find a good jointly-distributed assignments X_v for $\mathrm{val}(G)$. We will sample from $\Omega'_v = \{\omega \in \Omega' | f'_\omega(v, \beta) > 0\}$ for a unique β. To be more precise, we can sample X_v as follows: randomly permute Ω' and sample ω until we arrive at a f_ω such that there exists some unique β such that $f_\omega(v, \beta) = 1$ and assign v the answer β. Each player does, this obtaining a shared assignment for f in the symmetrized game. We now just have to bound the probability of success i.e. when (X_{v_1}, X_{v_2}) satisfies the symmetrized constraint $\mathcal{W} = \mathcal{V}_{sym}$. We lower bound this probability by also requiring $R(v_1) = R(v_2)$ where $R(v)$ is the number of times we have to sample until v gets assigned a label. The probability of this event over X is equivalent to the probability over ω that the constraint is satisfied conditioned on v_1 or v_2 receiving a label from f_ω since that would imply that f_ω gave the proper assignment for both v_1 and v_2 resulting in $R(v_1) = R(v_2)$. In other words we let $x_{\omega,v}$ be the value X_v takes on the partial assignment f_ω where we sample $\omega \sim \Omega'$ and we get the following

$$\begin{aligned}
\Pr_X[\mathcal{W}(X_{v_1}, X_{v_2}) = 1] &\geq \Pr_X[\mathcal{W}(X_{v_1}, X_{v_2}) = 1 \wedge R(v_1) = R(v_2)] \\
&= \Pr_\omega[\mathcal{W}(x_{\omega,v_1}, x_{\omega,v_2}) = 1 | f_\omega(v_1, x_{\omega,v_1}) \vee f_\omega(v_2, x_{\omega,v_2})] \\
&= \frac{\Pr_\omega[\mathcal{W}(x_{\omega,v_1}, x_{\omega,v_2}) = 1 \wedge (f_\omega(v_1, x_{\omega,v_1}) \vee f_\omega(v_2, x_{\omega,v_2}))]}{\Pr_\omega[f_\omega(v_1, x_{\omega,v_1}) \vee f_\omega(v_2, x_{\omega,v_2})]} \\
&= \frac{\mathbb{E}_\omega[\mathcal{W}(x_{\omega,v_1}, x_{\omega,v_2}) \min(f_\omega(v_1, x_{\omega,v_1}), f_\omega(v_2, x_{\omega,v_2}))]}{\mathbb{E}_\omega[\max(f_\omega(v_1, x_{\omega,v_1}), f_\omega(v_2, x_{\omega,v_2}))]}
\end{aligned}$$

The remainder of the analysis is omitted but follows from a lemma regarding random variables which we mention below.

Lemma 3.5. Let A, B, Z be jointly-distributed random variables such that A, B are nonnegative-valued and Z is 0/1-valued. Then, $\mathbb{E}\, Z \cdot \min A, B \geq (\frac{1-\gamma}{1+\gamma})\, \mathbb{E} \max\{A, B\}$ as long as $\mathbb{E}\, Z \cdot \min\{A, B\} \geq (1-\gamma')\, \mathbb{E}\, \frac{1}{2}(A+B)$.

4 Closing Remarks

4.1 Statistical vs. Graphical Approach

Rao also developed a result for projection games by fine tuning Holenstein's simplification, specifically the value is bounded by $(1-\epsilon/2)^{\frac{\epsilon n}{6\cdot(68)^2}}$ which actually removes the dependence on the answer length and this follows from focusing on the number of answers which are used frequently which he calls *heavy*. This is analogous to the idea of collision value which Dinur and Steurer introduce which isn't too surprising since projection games imply many answers α project onto a low number of β answers. However Raz's result falters when the number of repetitions is small e.g. $k \ll \frac{1}{\epsilon}$, and turns out to be weaker than the trivial bound $1-\epsilon$ while Dinur and Steurer's is tight. Does this imply the statistical approach is insignificant compared to the spectral graph theory approach? It is hard to say since the former is much easier to follow while the latter formalizes the notion of projection games quite nicely but it does not seem like it can be extended to general games. The statistical distribution approach allows for a bit of freedom in approximating the distributions involved in the games while Dinur and Steurer's analysis is very careful since we consider all finite sized measure spaces to find a good bound for ρ_G. The latter also focuses on the relationship between two arbitrary projection games in parallel rather than focusing on G^n and also relied on the trivial game T_v as a basis for amplification. In the end both rely on some modification game values, conditioning the distributions in the case of Holenstein, and the relaxed value val_+ in the case of Dinur and Steurer and so a correlated sampling was necessary to generate a good assignment for the original game. One may be able to take some definitions from Dinur and Steurer, for example the notion of a symmetrized game, and combine it with the statistical analysis of Rao to close the gap between the results.

4.2 Strong Parallel Repetition

For a while, researchers were wondering if parallel repetition can be strengthened further to $(1-\epsilon)^{\Omega(n/c)}$ where c is the answer length. However this turns out not to be true as Raz provides a counterexample through an odd-cycle XOR game[7]. The game involves a cycle of odd length $m \geq 3$ that the players are trying to show is 2-colorable. The referee asks each player the color of a common vertex with probability $\frac{1}{2}$, accepting if their answers match, and adjacent vertices with probability $\frac{1}{2}$, accepting if their answers differ. The value of a single game is $1 - 1/2m$ and Raz shows the repeated game has value $1 - (1/m) \cdot O(\sqrt{n})$. Dinur and Steurer show this bound is tight for projection games with few repetitions using their result. Raz uses a statistical approach similar to Holenstein/Rao and bounds the statistical distance between carefully selected edge distribution, one for each vertex. After he develops the closeness of these distributions, each player generates one based on their received vertices and use correlated sampling to show that the edge sets are the same and hence get a desired distribution on the vertex values. The fact that the statistical approach here gives the tightness to Dinur and Steurer's result via analytical methods is fairly interesting and suggests that such an approach may be able to be improved or generalized to give stronger parallel repetition results.

4.3 Open Problems

We have only touched the surface of parallel repetition but many generalizations are currently being worked on by many researchers in complexity theory. Two of the major ones include when the players are entangled (Quantum Parallel Repetition) and when there are more than 2 players (Multiplayer parallel repetition), both of which are open. Recent studies have shown that certain multi-player and quantum games classified as *anchoring* games satisfy some form of the parallel repetition theorem[1]. In addition there exists *anchoring transformations* which take a game G to an anchored game $G_{\perp}$. These transformations also rely on the graphical representation of games by adding anchored vertices and edges, and if any anchored edge is selected the game is automatically won. As a result the the winning probability is slightly amplified for a single game but still obeys exponential decay as the number of repeated games increase since $\perp$ disrupts Alice's and Bob's to use one round's results to influence another, the core to the general proof of parallel repetition. Extending parallel repetition to specific types of games allow us to extend its applications, most of which with deal with hardness of proof protocols. Researches are constantly reinventing the wheel, adding interesting features to these protocols but it is not certain whether parallel repetition is likely to hold or not on these protocols, and we may need more general techniques to find out.

References

[1] M. Bavarian, T. Vidick, H. Yuen, *Anchoring games for parallel repetition,* 2015.

[2] I. Dinur, D. Steurer, *Analytical Approach to Parallel Repetition,* 2014.

[3] J. Håstad, R. Pass, et al. *An Efficient Parallel Repetition Theorem,* Proceedings of the 7th International Conference on Theory of Cryptography, 2010.

[4] A. Rao, *Parallel Repetition in Projection Games and a Concentration Bound,* Institute of Advanced Study, 2009.

[5] A. Rao, *Information Theory in Computer Science Lecture 10,* University of Washington CSE533, 2010.

[6] R. Raz, *A Parallel Repetition Theorem,* SIAM J. Comput, 27(3), 763-803, 1998.

[7] R. Raz, *A Counterexample to Strong Parallel Repetition,* 2014.

18.405J / 6.841J Advanced Complexity Theory
Spring 2016