Parallel Repetition in Projection Games and a Concentration Bound

Anup Rao*
Institute for Advanced Study
arao@ias.edu

February 5, 2009

Abstract

In a two player game, a referee asks two cooperating players (who are not allowed to communicate) questions sampled from some distribution and decides whether they win or not based on some predicate of the questions and their answers. The parallel repetition of the game is the game in which the referee samples n independent pairs of questions and sends corresponding questions to the players simultaneously. If the players cannot win the original game with probability better than $(1-\epsilon)$, what's the best they can do in the repeated game?

We improve earlier results [Raz98, Hol07], which showed that the players cannot win all copies in the repeated game with probability better than $(1-\epsilon^3)^{\Omega(n/c)}$ (here c is the length of the answers in the game), in the following ways:

- We prove the bound $(1-\epsilon^2)^{\Omega(n)}$ as long as the game is a "projection game", the type of game most commonly used in hardness of approximation results. Our bound is independent of the answer length and has a better dependence on ϵ. By the recent work of Raz [Raz08], this bound is tight. A consequence of this bound is that the Unique Games Conjecture of Khot [Kho02] is equivalent to:

  Unique Games Conjecture There is an unbounded increasing function $f : \mathbb{R}^+ \to \mathbb{R}^+$ such that for every $\epsilon > 0$, there exists an alphabet size $M(\epsilon)$ for which it is NP-hard to distinguish a Unique Game with alphabet size M in which a $1-\epsilon^2$ fraction of the constraints can be satisfied from one in which a $1-\epsilon f(1/\epsilon)$ fraction of the constraints can be satisfied.

- We prove a concentration bound for parallel repetition (of general games) showing that for any constant $0 < \delta < \epsilon$, the probability that the players win a $(1-\epsilon+\delta)$ fraction of the games in the parallel repetition is at most $\exp\left(-\Omega(\delta^4 n/c)\right)$. An application of this is in testing Bell Inequalities. Our result implies that the parallel repetition of the CHSH game can be used to get an experiment that has a very large classical versus quantum gap.

Keywords: Parallel Repetition, Hardness of Approximation, Bell Inequalities, CHSH Game

*Supported by the National Science Foundation under agreement No. CCR-0324906.

1 Introduction

In this paper we study the following type of two player game ($\mathcal{G}$) — the game is defined by a distribution on questions (X, Y) and a predicate V. A referee administers the game by sampling two questions (x, y) and sending one question to each of the players. The players respond with answers $(a(x), b(y))$ and they win if and only if the predicate $V(x, y, a, b)$ is satisfied. The players are not allowed to communicate during the game though they may use shared randomness. The *value* of the game (usually denoted by $(1 - \epsilon)$ in this paper) is the maximum probability of success that the players can achieve. Note that any strategy for the players that involves the use of shared randomness can be matched by a deterministic strategy that achieves the same probability of success, simply by fixing the shared randomness in the best possible way.

These games are interesting for several reasons. Any one round two prover interactive proof (as introduced by [BGKW88]) can be modeled as a game between the verifier and the two players. This proof system turns out to be powerful enough to capture all of non-deterministic exponential time (NEXP), with an exponentially small error. Such games also arise in cryptographic applications [BGKW88, BGKW89, DFK+92, LS95], hardness of approximation results [FGL+91, ALM+98, FL92, LY93], and have been used to prove direct product theorems for communication complexity [PRW97].

Given any game $\mathcal{G}$, the n-fold parallel repetition of the game $\mathcal{G}^n$ is the natural game in which the referee samples n independent questions $(X_1, Y_1), \ldots, (X_n, Y_n)$, each distributed according to (X, Y), and sends all the x questions to the first player and all the y questions to the second player. The players then each respond with n answers, and the referee decides that they win if and only if they win in each of the n coordinates. For each player, the answer in the i'th coordinate may depend on the question asked in some other coordinate. This paper is about bounding the probability that the players can win the n-fold parallel repetition.

1.1 Previous Work

The first bound on the value of repeated games was obtained by Verbitsky [Ver94] who showed that the value of the game $\mathcal{G}^n$ must tend to 0 as n tends to infinity. This was followed by a much stronger bound due to Raz, which involved the *answer length*, of the game. The answer length is c if the number of answers that the players can give is bounded by 2^c. Raz proved that:

Theorem 1.1 (Raz [Raz98])**.** *There is a universal constant $\alpha > 0$ and a function $0 < \epsilon'(\epsilon) < 1$ such that the value of $\mathcal{G}^n$ is at most $(1 - \epsilon')^{\alpha n/c}$.*

The dependence on the answer length c was shown to be almost optimal by Feige and Verbitsky [FV02]. Raz's proof was recently simplified by Holenstein who gave a more explicit bound:

Theorem 1.2 (Holenstein [Hol07])**.** *There is a universal constant $\alpha > 0$ such that the value of $\mathcal{G}^n$ is at most $(1 - \epsilon)^{\alpha\epsilon^2 n/c}$.*

Proving our results requires us to revisit Holenstein's proof and modify it at certain points, so we include a proof of his result as well.

1.2 Projection/Unique Games

A *projection game* is a game in which the predicate V has a special kind of structure — every pair (x, y) defines a function f_{xy} and the predicate V is satisfied exactly when $f_{xy}(a) = b$. If the game is

such that f_{xy} is a permutation for every xy, then the game is called a *unique game*, since every answer of one player induces a unique answer for the other player.

Both of these types of games have played an important role in many hardness of approximation results. Most hardness of approximation results are proved by giving a reduction to the Label Cover problem. An instance of this problem is a bipartite graph where every edge (x, y) is associated with a function f_{xy} and the problem is to estimate the maximum number of edges that can be satisfied by any assignment of values $a(x), b(y)$ to the vertices in the graph (an edge xy is satisfied if $f_{xy}(a(x) = b(y))$). Every such instance L is associated with a projection game $\mathcal{G}_L$ in the natural way (the referee picks a random edge, asks the players for assignments to the vertices and checks to see that the projection is satisfied), and the problem of finding the best assignment to L is the same as the problem of finding the best strategy for $\mathcal{G}_L$. Similarly, the parallel repetition of the game $\mathcal{G}_L^n$ is associated with another instance of Label Cover L^n, where again the value of $\mathcal{G}_L^n$ is the maximum fraction of edges of L^n that can be satisfied by any labeling. Further, the property of being a unique or projection game is preserved under parallel repetition.

The PCP theorem [AS98, ALM+98] shows that there exists a constant $\epsilon_0 > 0$ and an alphabet size 2^c, such that it is NP-hard to distinguish a Label Cover instance with alphabet size 2^c that has value 1 from an instance that has value $1 - \epsilon_0$. Combining this with the parallel repetition theorem of Raz mentioned above, we get that for any constant $1 > \epsilon > 0$, it is in fact NP-hard to distinguish instances of Label Cover with value 1 from instances with value ϵ. This follows just by taking the instance L obtained from the PCP theorem and using Raz's theorem to bound the value of L^n. If L has value 1, then L^n clearly still has value 1. On the other hand, if L has value at most $1 - \epsilon_0$, n can be chosen to be large enough so that L^n has value at most ϵ.

For Unique Games, it is impossible to get such a 1 vs $(1-\epsilon)$ hardness result (unless P=NP) with a small alphabet, since there is a trivial algorithm that can check if a Unique Game instance has value 1 or not — simply try all assignments to a single vertex v in the graph. Every such assignment induces unique assignments to all other vertices in the graph if a consistent assignment exists. Still, we may hope that the following conjecture (due to Khot [Kho02]) is true:

Conjecture 1.3 (Unique Games Conjecture)**.** *For every ϵ, there exists an answer length $c(\epsilon)$ for which it is NP-hard to distinguish instances of Unique Games with answer length c that have value at least $1 - \epsilon$ from instances that have value at most ϵ.*

Several tight or almost tight hardness results have been proved assuming the Unique Games Conjecture, including for Max 2-Lin [Kho02], Vertex Cover [KR03], Max-Cut [Kho02, KKMO04, MOO05], Approximate Coloring [DMR06], Sparsest Cut [CKK+06, KV05] and Max 2-Sat [Aus07]. Thus the question of whether or not the conjecture is true is of considerable interest. On the negative (algorithmic) side, approximation algorithms [Tre05, CMM06, CMM06, GT06] have been designed to approximate the value of a Unique Game. For example, given a unique games instance with value $1 - \epsilon$, an algorithm due to Charikar, Makarychev and Makarychev [CMM06] can find an assignment with value $1-O(\sqrt{\epsilon c})$. This implies that the answer length $c(\epsilon)$ in the Unique Games Conjecture *must* be larger than $\Omega(1/\epsilon)$ if the conjecture is to hold.

On the positive side, we might have hoped that we could use the parallel repetition thereoms of Raz or Holenstein to reduce the task of proving the conjecture to the task of proving it for a much smaller gap, just as we did above for the case of Label Cover, and then try and prove the conjecture for that small gap. However, the bounds of Raz and Holenstein are problematic for this purpose. If $\epsilon > \delta$ and we try to apply Holenstein's theorem to amplify a gap of $(1 - \delta)$ vs $1 - \epsilon$, n repetitions give us the gap $(1 - \delta)^n$ vs $(1 - \epsilon^3)^{\alpha n/c}$, which is the right kind of gap only if $\delta \ll \alpha\epsilon^3/c$. A conjecture

with this kind of gap cannot hold, since the algorithm of Charikar et al. shows that if δ, ϵ, c satisfy this constraint, we can distinguish $(1-\delta)$ instances from $(1-\epsilon)$ instances in polynomial time.

Another way in which a strong parallel repetition might have helped to resolve the Unique Games Conjecture is by using a bound on the value of the parallel repetition to show that the conjecture is equivalent to the hardness of approximating other problems. A good candidate for such a problem is Max-Cut, every instance of which is a special kind of Unique Game. The results of Khot et al. [Kho02, KKMO04, MOO05] show that the Unique Games Conjecture implies that that Goemans-Williamson algorithm for Max-Cut [GW95], which can distinguish instances with value $1-\epsilon^2$ from those with value $1-\epsilon$, is essentially optimal. It is conceivable that the Unique Games Conjecture is actually equivalent to the optimality of this algorithm. If we could prove a bound of the form $(1-\epsilon^{2^-})^{\Omega(n)}$, we could use an approximation algorithm for Unique Games to get an algorithm for Max-Cut, just by running the algorithm on the parallel repetition of the Max-Cut instance. Unfortunately, Raz recently discovered [Raz08] that such a bound is impossible, even if we restrict our attention to games that come from instances of MAX-Cut (his counterexample is the MAX-Cut game played on an odd cycle). He proved that the parallel repetition of a MAX-Cut game on an odd cycle (which is a Unique Game) has value at least $(1-\epsilon^2)^{O(n)}$, killing the hopes of getting such an equivalence via a stronger parallel repetition bound.

1.3 Parallel Repetition and Bell Inequalities

Two player games also show up in the context of testing so called *Bell Inequalities* [Bel64] to confirm the existence of quantum entanglement. The idea is to consider games where two players who have access to entangled qubits can achieve a much higher success probability than two classical players can. Perhaps the most famous example of a game where such a gap exists is the CHSH game [CHSH69]. Here the verifier sends the players random bits (x, y) and receives one bit answers $(a(x), b(y))$. The players win when $a \oplus b = x \wedge y$. Two classical players cannot win with probability better than 0.75, but it can be shown that two players sharing entangled qubits can win with probability close to 0.85.

For the purpose of testing Bell Inequalities, it is important to be able to come up with games that have a big gap between the success probability of classical players and the success probability of entangled players, and this seems to be an issue that has warranted a significant amount of effort [BCH+02, Gil03].

This motivates proving a *concentration bound* for parallel repetition. If we could prove that two players cannot hope to win more than the expected fraction of games in the parallel repetition, we would get a simple way to construct games with a large quantum vs classical gap. We can take the CHSH game (any game with a small classical vs quantum gap would work) and consider its n-fold parallel repetition. We say that the players win the repeated game as long as they win in 0.8 fraction of the coordinates. The concentration bound would imply that if the players were classical, they can win this game with a very small probability. On the other hand, the Chernoff bound shows that the obvious quantum strategy (play each game in the repetition independently) is sure to win the game with all but exponentially small probability.

1.4 Our Results

- We prove an essentially tight bound on the value of the parallel repetition in the case that the original game is a projection game.

  Theorem 1.4 (Parallel Repetition in Projection Games)**.** *There is a universal constant $\alpha > 0$ such that if $\mathcal{G}$ is a projection game with value at most $1-\epsilon$, the value of $\mathcal{G}^n$ is at most $(1-\epsilon)^{\alpha\epsilon n}$.*

This improves on the earlier bounds in two ways: the dependence on ϵ is better ($(1-\epsilon)^{\alpha\epsilon n} \approx (1-\epsilon^2)^{\alpha n}$), and the answer length does not even show up in the bound. Of course, every unique game is also a projection game, so this theorem can be used to amplify the gap between unique games instances. Working out the parameters gives that the Unique Games Conjecture is equivalent to the following statement:

Conjecture 1.5 (Unique Games Conjecture Restated)**.** *There exists an unbounded increasing function $f : \mathbb{R}^+ \to \mathbb{R}^+$ such that for every $\epsilon > 0$, there is an answer length $c(\epsilon)$ for which it is NP-hard to distinguish instances of unique games with answer length c that have value $1-\epsilon^2$ from instances that have value $1-\epsilon f(1/\epsilon)$.*

As we discussed above, the work of Raz [Raz08] shows that our bound is tight upto the constant α.

- We prove a concentration bound for parallel repetition:

  Theorem 1.6 (Concentration in General Games)**.** *There is a universal $\alpha > 0$ such that if $\mathcal{G}$ is a game with value $1-\epsilon$ and answer length c, for every $\epsilon \geq \delta > 0$, the probability that the players win more than a $1-\epsilon+\delta$ fraction of the games in the n-fold parallel repetition is bounded by $\exp\left(-\frac{\alpha\delta^4 n}{c}\right)$.*

  Following our discussion, this bound shows that the parallel repetition of the CHSH game gives a game with a large classical vs quantum gap.

1.5 Techniques

Our proofs build on the work of Raz and Holenstein. In this section we shall be vague (and slightly inaccurate) in order to convey what is new about our work without revealing too many technical details.

Fix a strategy for $\mathcal{G}^n$. We use the notation $X^n = X_1, \ldots, X_n$ and $Y^n = Y_1, \ldots, Y_n$ to denote the questions that are asked to the players in $\mathcal{G}^n$. It turns out that the heart of all the earlier proofs (and our own) is a lemma of the following type:

Informal Lemma 1.7. *Let $S \subset [n]$ be any set of small size k and W_S denote the event that the players win the games corresponding to the coordinates in S. Then, if $\Pr[W_S]$ is* large enough *there exists an index $i \notin S$ such that the probability that the players win the i'th coordinate conditioned on W_S is at most $1-\epsilon/2$.*

Here we need $\Pr[W_S]$ to be larger than some function of ϵ, n, k and the answer length c. Once we have this kind of lemma, it is not too hard to show that the players cannot win $\mathcal{G}^n$ with a high probability, and we leave this to the formal parts of the paper.

The lemma is proved via a reduction — we can show that if the lemma is false, i.e. if there exists a small set S and a dense event W_S for which the lemma is false, we can find a strategy for $\mathcal{G}$ that wins with probability larger than $1-\epsilon$, which is a contradiction. Suppose there exists such a set S for which W_S is dense. Then the players decide on an index i ahead of time. When asked the questions (X, Y), the players place these questions in the i'th coordinate and use shared randomness to generate $n-1$ other pairs of questions (X_j, Y_j) such the joint distribution of the questions they end up with is $\epsilon/2$ close to $\left(X^n, Y^n \middle| W_S\right)$ in statistical distance. Since the lemma is assumed to be false, the players

can then use the i'th coordinate answers dictated by the strategy for $\mathcal{G}^n$ to win $\mathcal{G}$ with probability more than $1 - \epsilon$.

The questions are actually generated in two steps. In the first step, the players *simultaneously* sample two random variables R, A, i.e. they end up with the same sample for these random variables with high probability. The random variable A is just the answers of the first player in the coordinates in S. The random variable R contains at least one question from every pair (X_j, Y_j), and both questions from the pairs corresponding to the coordinates in S. These properties allow us to argue that for every r, a, $\left(X^n, Y^n \middle| (R, A) = (r, a) \wedge W_S\right)$ is a product distribution. This means that once the players have agreed on the sample for R, A, they can use independent randomness to sample the rest of the questions conditioned on the information they have, and end up with a distribution on questions that is close to $\left(X^n, Y^n \middle| W_S\right)$.

It turns out that $\Pr[W_S \wedge A = a | R = r]$ needs to be *large enough* for typical fixings of $R = r$ for this argument to go through. Raz and Holenstein argue that this quantity is large, just by counting. They argue that if the answer length is c bits, $\Pr[W_S \wedge (R, A) = (r, a)] / \Pr[W_S \wedge R = r]$ is typically at least 2^{-ck}, since there are at most 2^{ck} possible ways to set the random variable A if the answer length is c. In our work, we get a stronger bound by observing that in the case of a projection game, the players cannot be using *all* of their answers equally often.

For simplicity, let us assume that the game is unique. Then note that for every fixing of $R = r$, the bijection between the answers of the players in the coordinates of S is determined, but the answers are now two independent random variables. It is not too hard to show that if two independent random variables satisfy some bijection with probability γ, there must exist a set of size $100/\gamma$ such that the probability that the bijection is satisfied and the first random variable does not land in this set is less than $\gamma/100$ (simply take the the set to be the elements of weight at least $\gamma/100$). The argument also works in the case that the constraints are projections instead of bijections.

So we can argue that for every fixing of $R = r$, there is a small set of *heavy answers* that the players must be using. This argument lets us get a lowerbound on $\Pr[W_S \wedge (R, A) = (r, a)]$ that is independent of the answer length.

To prove the concentration bound, we first observe that the lemma above can be generalized slightly in the following way:

Informal Lemma 1.8. *Let $S \subset [n]$ be any set of small size k and E be* any *event that is determined by what happens in the games of S. Then, if* $\Pr[E]$ *is* large enough, most *indices $i \notin S$ are such that the probability that the players win the i'th coordinate conditioned on E is at most* $1 - \epsilon/2$.

Once we have this lemma, we can show that if the referee samples a small fraction of the coordinates uniformly at random and checks that the players have won in those coordinates, his count of how many games the players have won in the random sample behaves like a supermartingale; conditioned on the result of his sampling so far, the outcome at the next random coordinate is biased towards losing. This allows us to bound the probability that the referee sees a larger fraction of wins than he should. On the other hand, the Chernoff bound gives that with high probability, the referee's experiment gives a good estimate for the fraction of games that the players won. These arguments allow us to bound the probability that the players win a large fraction of the games.

2 Preliminaries

2.1 Notation

We use calligraphic letters to denote sets, capital letters to denote random variables and small letters to denote instantiations of random variables/elements of sets. We shall use the same letter to denote objects of this type that are related to each other. For example, we shall use X to denote a random variable taking values in the set $\mathcal{X}$ and x to denote an instantiation of that random variable.

In this paper we shall often need to start with some probability space and modify it in certain ways. We explain our notation with the help of some examples. If A, B, C are random variables in some probability space taking values in $\mathcal{A}, \mathcal{B}, \mathcal{C}$, then:

- $A'B'C' \stackrel{def}{=} \{A\}\{B\}\{C\}$ defines a new probability space in which A', B', C' take values in $\mathcal{A}, \mathcal{B}, \mathcal{C}$ such that
$$\Pr[A' = a \wedge B' = b \wedge C' = c] \stackrel{def}{=} \Pr[A = a] \Pr[B = b] \Pr[C = c]$$
- $A'B'C' \stackrel{def}{=} \{AB\}\{C\}$ means
$$\Pr[A' = a \wedge B' = b \wedge C' = c] \stackrel{def}{=} \Pr[A = a \wedge B = b] \Pr[C = c]$$
- $A'B'C' \stackrel{def}{=} \{AB\}\left\{C\middle|B\right\}$ means
$$\Pr[A' = a \wedge B' = b \wedge C' = c] \stackrel{def}{=} \Pr[A = a \wedge B = b] \Pr[C = c|B = b]$$
- Let $\tilde{A}$ be a random variable taking values in $\mathsf{supp}(A)$. Then $A', B' \stackrel{def}{=} \left\{\tilde{A}\right\}\left\{B\middle|\tilde{A}\right\}$ means
$$\Pr[A' = a \wedge B' = b] \stackrel{def}{=} \Pr[\tilde{A} = a] \Pr[B = b|A = a]$$

2.2 Statistical Distance

Sometimes the distributions we get are not exactly the distributions we want, but they may be *close* enough. The measure of *closeness* we will use is this one:

Definition 2.1. Let D and F be two random variables taking values in a set $\mathcal{S}$. Their *statistical distance* is
$$|D - F| \stackrel{def}{=} \max_{\mathcal{T} \subseteq \mathcal{S}} (|\Pr[D \in \mathcal{T}] - \Pr[F \in \mathcal{T}]|) = \frac{1}{2} \sum_{s \in \mathcal{S}} |\Pr[D = s] - \Pr[F = s]|$$

If $|D - F| \leq \epsilon$ we shall say that D is ϵ-*close* to F. We shall also use the notation $D \stackrel{\epsilon}{\approx} F$ to mean D is ϵ-close to F.

Proposition 2.2. *Let D and F be any two random variables over a set $\mathcal{S}$ s.t. $|\mathsf{P}_D - \mathsf{P}_F| \leq \epsilon$. Let g be any function on $\mathcal{S}$. Then $|g(D) - g(F)| \leq \epsilon$.*

Proposition 2.3 (Triangle Inequality)**.** *Let A, B, C be random variables over $\mathcal{S}$, with $A \stackrel{\epsilon_1}{\approx} B \stackrel{\epsilon_2}{\approx} C$. Then $A \stackrel{\epsilon_1+\epsilon_2}{\approx} C$.*

Proposition 2.4 (Conditioning Close Distributions)**.** *Let A, B be two random variables and let E_1, E_2 be two events with $\Pr[E_i] = \mu$. Then $|A|E_1 - B|E_2| \leq |A - B|/\mu$.*

Proposition 2.5. *Let A, A' be two random variables over $\mathcal{A}$ in the same probability space such that $\Pr[A \neq A'] \leq \epsilon$. Then $|A - A'| \leq \epsilon$.*

Proof. Let $\mathcal{S} \subset \mathcal{A}$ be any set. Then by the union bound we get $\Pr[A \in \mathcal{S}] \leq \Pr[A' \in \mathcal{S}] + \Pr[A \neq A']$, which clearly implies the proposition. □

2.3 Games

In this paper, a game is defined by a distribution (X, Y) on a set of questions, $\mathcal{X} \times \mathcal{Y}$, a set of possible answers $\mathcal{A} \times \mathcal{B}$ and a predicate $V : \mathcal{X} \times \mathcal{Y} \times \mathcal{A} \times \mathcal{B}$. A strategy for the game is a pair of functions $a : \mathcal{X} \to \mathcal{A}$ and $b : \mathcal{Y} \to \mathcal{B}$. The value of the game is the maximum of $\Pr_{X,Y}[V(X, Y, a(X), b(Y)]$, over all choices of strategies $a(\cdot), b(\cdot)$.

We call a game a *projection* game if there exists a family of functions $f_{x,y}$ indexed by $\mathcal{X} \times \mathcal{Y}$ such that $V(x, y, a, b)$ is equivalent to $f_{x,y}(b) = a$.

A game is called *unique* if it is a projection game with the additional property that all function $f_{x,y}$ are bijections.

The *answer length* of a game is the quantity $\log |\mathcal{A}| + \log |\mathcal{B}|$.

Given a game $\mathcal{G}^n$ the parallel repetition of the game is the game with distribution on questions obtained by taking n independent samples $(X_1, Y_1) \cdots (X_n, Y_n)$. A strategy for the new game is specified by two functions $a : \mathcal{X}^n \to \mathcal{A}^n$ and $b : \mathcal{Y}^n \to \mathcal{B}^n$. When the game is played, the referee samples n independent pairs of questions as above and sends one question from each pair to each of the players. The players respond with n answers each. The referee then checks that the players win by checking the AND of the predicate in the original game in each of the n copies. Thus the value of the game is the maximum of $\Pr[V(X_1, Y_1, a_1(X_1), b_1(Y_1) \wedge \cdots \wedge V(X_n, Y_n, a_n(X_n), b_n(Y_n)]$, over all choices of strategies $a(\cdot), b(\cdot)$.

3 Main Theorems

In this section, we prove our main theorems, assuming two lemmas which we prove in a later section. Fix an optimal strategy for the repeated game. For any set $S \subset [n]$, we let W_S denote the event that the players win all the games in coordinates included in the set S.

Lemma 3.1 (Main Lemma for General Games)**.** *Let $S \subset [n]$ be of size k. If the game $\mathcal{G}$ is such that one player gives answers from a set of size 2^c and $\Pr[W_S] \geq 2^{-\frac{\epsilon^2(n-k)}{34^2}+kc}$, then $\mathbb{E}_{i \notin S}\left[\Pr[W_{\{i\}}|W_S]\right] \leq 1 - \epsilon/2$.*

Lemma 3.2 (Main Lemma for Projection Games)**.** *Let $S \subset [n]$ be of size k. If $\mathcal{G}$ is a projection game with $n - k \geq \frac{5(68)^2 \log(4/\epsilon)}{2\epsilon^2}$ and $\Pr[W_S] \geq 2^{-\frac{\epsilon^2(n-k)}{5(68)^2}}$, then $\mathbb{E}_{i \notin S}\left[\Pr[W_{\{i\}}|W_S]\right] \leq 1 - \epsilon/2$.*

We can now prove the theorems assuming these lemmas.

Theorem 3.3 (Main Theorem for General Games)**.** *If $\mathcal{G}$ is a game with value $(1 - \epsilon)$ and answer length c bits, then the n-fold parallel repetition has value at most $(1 - \epsilon/2)^{\frac{\epsilon^2 n}{35^2+34^2 c}}$.*

Proof. We have chosen the exponent $t = \frac{\epsilon^2 n}{35^2+34^2 c}$ in the theorem so that it satisfies $\epsilon t \leq \frac{\epsilon^2(n-t)}{34^2} - tc$. Suppose for the sake of contradiction that $\Pr[W_{[n]}] > (1-\epsilon/2)^t$. Then for every set $H \subset [n]$, $\Pr[W_H] > (1-\epsilon/2)^t$. Let k be the smallest number for which every set $H \subset [n]$ of size $k+1$ satisfies $\Pr[W_H] > (1-\epsilon/2)^{k+1}$.

Our assumption means that $k+1 \leq t \Rightarrow k < t$. By the choice of k, there must exist a set $S \subset [n]$ of size k for which $\Pr[W_S] \leq (1-\epsilon/2)^k$. Then

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

Where here we used the inequality $(1-\epsilon/2) \geq 2^{-\epsilon}$ for $\epsilon \in [0,1]$, the fact that $k < t$ and the bound on ϵt. We can now apply Lemma 3.1 to show that there exists an index i with

$$\Pr[W_{S\cup\{i\}}] = \Pr[W_{\{i\}}|W_S]\Pr[W_S] \leq (1-\epsilon+17\epsilon/34)(1-\epsilon/2)^k = (1-\epsilon/2)^{k+1}$$

contradicting our choice of k. ∎

Next we give the theorem for projection games.

Theorem 3.4 (Main Theorem for Projection Games)**.** *If $\mathcal{G}$ is a projection game with value $(1-\epsilon)$, the n-fold parallel repetition has value at most $(1-\epsilon/2)^{\frac{\epsilon n}{6(68)^2}}$.*

Proof. First we prove the theorem for the case that $n \geq \frac{6(68)^2\log(4/\epsilon)}{2\epsilon^2}$. The proof is very similar to the proof of the previous case. We have engineered the exponent $t = \frac{\epsilon n}{6(68)^2}$ to satisfy $\epsilon t \leq \frac{\epsilon^2(n-t)}{5(68)^2}$. The bound on n was chosen to ensure $n - t \geq \frac{5(68)^2\log(4/\epsilon)}{2\epsilon^2}$.

Suppose for the sake of contradiction that there is a strategy for which $\Pr[W_{[n]}] > (1-\epsilon/2)^t$, which implies that for every set $H \subset [n]$, $\Pr[W_H] > (1-\epsilon/2)^t$. Let k be the smallest number for which every set $H \subset [n]$ of size $k+1$ satisfies $\Pr[W_H] > (1-\epsilon/2)^{k+1}$.

Then we get that $k+1 \leq t \Rightarrow k < t$. By our choice of k, there must exist a set $S \subset [n]$ of size k for which $\Pr[W_S] \leq (1-\epsilon/2)^k$. Then

$$\Pr[W_S] \geq (1-\epsilon/2)^t \geq 2^{-\epsilon t} \geq 2^{-\frac{\epsilon^2(n-t)}{5(68)^2}} \geq 2^{-\frac{\epsilon^2(n-k)}{5(68)^2}}$$

Where here we used the inequality $(1-\epsilon/2) \geq 2^{-\epsilon}$ for $\epsilon \in [0,1]$ and the fact that $k < t$. We also have that $n - k \geq n - t \geq \frac{5(68)^2\log(4/\epsilon)}{2\epsilon^2}$. We can now apply Lemma 3.2 to show that there exists an index i with

$$\Pr[W_{S\cup\{i\}}] = \Pr[W_{\{i\}}|W_S]\Pr[W_S] \leq (1-\epsilon/2)(1-\epsilon/2)^k = (1-\epsilon/2)^{k+1}$$

contradicting our choice of k.

Next note that if the above bound holds for large n, it must hold for small n[1]. Specifically, if there is some strategy for the players that achieves a probability of success greater than $(1-\epsilon/2)^{\frac{\epsilon n}{6(68)^2}}$ in the n-fold repetition, then for every r, simply repeating this strategy on disjoint coordinates gives a strategy for the nr-fold repetition with probability of success of $(1-\epsilon/2)^{\frac{\epsilon n r}{6(68)^2}}$. Setting r to be large enough contradicts the previous case. ∎

[1]This argument was suggested by an anonymous referee

4 Sampling From Close Distributions

Fix a set $\mathcal{A}$. A variant of the following lemma was proved by Holenstein [Hol07]. The proof we sketch here is due to Boaz Barak.

Lemma 4.1 (Sampling similar distributions [Hol07])**.** *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 $|A_l - A_i| \leq \epsilon_i$ for every $i \in [l]$, the players can use shared randomness to sample $B_1, \ldots, B_l$ with the property that:*

- *For every i, B_i has the same distribution as A_i.*
- *For every $i < l-1$, $\Pr[B_l \neq B_i] \leq 2\epsilon_i$.*
- $\Pr[\textit{all samples are the same}] \geq 1 - 2\sum_{i=1}^{l-1} \epsilon_i$

Proof Sketch: First note that the last guarantee follows from the second guarantee and the union bound.

To prove the first two guarantees, let us first consider the case that the A_i's are promised to be uniform over (possibly different) subsets of $\mathcal{A}$. In this case the protocol for the players is simple: the shared randomness is interpreted as a permutation of the universe $\mathcal{A}$. Each player then samples the first element of the permutation that lies in the support of her distribution. The lemma is then easily seen to be true.

To handle the general case, identify each distribution A_i with the uniform distribution on the set $\cup_{a \in \mathcal{A}} \{a\} \times [0, \Pr[A_i' = a]]$, which is a subset of $\mathcal{A} \times [0,1]$. Then by tiling the set $\mathcal{A} \times [0,1]$ with a fine enough grid, we can interpret the shared randomness as a permutation of the parts of this grid to get a protocol that is arbitrarily close to getting the bounds promised above. □

5 Conditioning Product Distributions

Lemma 5.1. *Let A, B be random variables in some probability space. Let A' be another random variable such that $|A - A'| \leq \epsilon$. Then $\left|\{AB\} - \{A'\}\left\{B\middle|A'\right\}\right| \leq \epsilon$.*

We shall need a basic definition:

Definition 5.2 (Informational Divergence)**.** Given two random variables U, V taking values in the same set $\mathcal{U}$, we define the informational divergence

$$\mathsf{D}\left(U\middle||V\right) \stackrel{def}{=} \sum_{u \in \mathcal{U}} \Pr[U = u] \log\left(\frac{\Pr[U=u]}{\Pr[V=u]}\right)$$

where we adopt the convention that $0 \log 0 = 0$. If there exists a $u \in \mathcal{U}$ for which $\Pr[V = u] = 0$ but $\Pr[U = u] \neq 0$, we say that that $\mathsf{D}\left(U\middle||V\right) = \infty$.

The following are standard facts about informational divergence:

Fact 5.3. $\mathsf{D}\left(V\middle||U\right) \geq |U - V|^2$

Fact 5.4. *If V is a random variable, E is any event and $\tilde{V} \stackrel{def}{=} V|E$, in the same space with $\Pr[E] = 2^{-d}$, then $\mathsf{D}\left(\tilde{V}\middle||V\right) \leq d$.*

Proof.

$$\begin{aligned}
&\mathsf{D}\left(\tilde{V}\middle|\middle|V\right)\\
&=\sum_{v\in\mathcal{V}}\Pr[V=v|E]\log\left(\frac{\Pr[V=v|E]}{\Pr[V=v]}\right)\\
&=\sum_{v\in\mathcal{V}}\Pr[V=v|E]\log\left(\frac{\Pr[E|V=v]}{\Pr[E]}\right)\\
&=\log(1/\Pr[E])+\sum_{v\in\mathcal{V}}\Pr[V=v|E]\log(\Pr[E|V=v])\\
&\leq\log(1/\Pr[E]) && \text{since every term in the sum is at most 0}
\end{aligned}$$

□

Fact 5.5. *If $U_1,\ldots,U_n$ are independent random variables and $V_1,\ldots,V_n$ are other random variables,*

$$\sum_{i=1}^{n}\mathsf{D}\left(V_i\middle|\middle|U_i\right)\leq\mathsf{D}\left(V_1\ldots V_n\middle|\middle|U_1\ldots U_n\right)$$

A key part of the proof will be showing that if we condition a product distribution on an event whose probability is not too low, there must be some coordinate which remains distributed how it was before the conditioning.

Lemma 5.6 ([Raz98])**.** *Let $U_1,U_2,\ldots,U_n$ be independent random variables. Suppose E is any event in the same probability space such that $\Pr[E]=2^{-d}$, then*

$$\mathop{\mathbb{E}}_{i\in[n]}\left[\left|\{U_i\}-\left\{U_i\middle|E\right\}\right|\right]\leq\sqrt{\frac{d}{n}}$$

Proof.

$$\begin{aligned}
&\mathop{\mathbb{E}}_{i\in[n]}\left[\left|\{U_i\}-\left\{U_i\middle|E\right\}\right|\right]^2\\
&\leq\mathop{\mathbb{E}}_{i\in[n]}\left[\left|\{U_i\}-\left\{U_i\middle|E\right\}\right|^2\right] && \text{by convexity of the square function}\\
&\leq\mathop{\mathbb{E}}_{i\in[n]}\left[\mathsf{D}\left(\left\{U_i\middle|E\right\}\middle|\middle|\{U_i\}\right)\right] && \text{by Fact 5.3}\\
&\leq\frac{1}{n}\mathsf{D}\left(\left(U_1U_2\ldots U_n\middle|E\right)\middle|\middle|U_1\ldots U_n\right) && \text{by Fact 5.5}\\
&\leq\frac{d}{n} && \text{by Fact 5.4}
\end{aligned}$$

□

Next, we show that all of the above still holds if in addition to dense event, we condition on the value of some random variable with small support, and the variables are only independent in convex combination:

Corollary 5.7. *Let $R,U_1,U_2,\ldots,U_n,A$ be random variables and E be an event with $\Pr[E]=2^{-d}$ such that*

- *For every r, $U_1, \ldots, U_n$ are independent conditioned on the event $R = r$.*
- *For every r, $|\mathsf{supp}\left(A \middle| E \wedge (R = r)\right)| \leq 2^h$*

Then,

$$\mathop{\mathbb{E}}_{i \in [n]} \left[\left| \left\{ RA \middle| E \right\} \left\{ U_i \middle| R \right\} - \left\{ RAU_i \middle| E \right\} \right| \right] \leq \sqrt{\frac{d+h}{n}}$$

Proof.

$$\begin{aligned}
&\mathop{\mathbb{E}}_{i \in [n]} \left[\left| \left\{ RA \middle| E \right\} \left\{ U_i \middle| R \right\} - \left\{ RAU_i \middle| E \right\} \right| \right]^2 \\
&= \mathop{\mathbb{E}}_{\substack{i \in [n] \\ (a,r) \leftarrow (AR|E)}} \left[\left| \left\{ U_i \middle| R = r \right\} - \left\{ U_i \middle| E \wedge (A,R) = (a,r) \right\} \right| \right]^2 \\
&\leq \mathop{\mathbb{E}}_{(a,r) \leftarrow (AR|E)} \left[\mathop{\mathbb{E}}_{i \in [n]} \left[\left| \left\{ U_i \middle| R = r \right\} - \left\{ U_i \middle| E \wedge (A,R) = (a,r) \right\} \right| \right]^2 \right] && \text{by convexity} \\
&\leq \mathop{\mathbb{E}}_{(a,r) \leftarrow (AR|E)} \left[\frac{\log(1/\Pr[E \wedge A = a | R, r])}{n} \right] && \text{by Lemma 5.6} \\
&\leq (1/n) \log \left(\mathop{\mathbb{E}}_{r \leftarrow (R|E)} \left[\sum_{a \in \mathsf{supp}(A|E \wedge R = r)} \frac{\Pr[A = a | E \wedge (R = r)]}{\Pr[E \wedge A = a | R = r]} \right] \right) && \text{by concavity of log} \\
&= (1/n) \log \left(\mathop{\mathbb{E}}_{r \leftarrow (R|E)} \left[\sum_{a \in \mathsf{supp}(A|E \wedge R = r)} \frac{\Pr[R = r]}{\Pr[E \wedge R = r]} \right] \right) \\
&\leq (1/n) \log \left(2^h \mathop{\mathbb{E}}_{r \leftarrow (R|E)} \left[\frac{\Pr[R = r]}{\Pr[E \wedge R = r]} \right] \right) \\
&= (1/n) \log \left(2^h \mathop{\mathbb{E}}_{r \leftarrow (R|E)} \left[\frac{\Pr[R = r]}{\Pr[E] \Pr[R = r | E]} \right] \right) \\
&= (1/n) \log \left(2^h \sum_{r \in \mathcal{R}} \frac{\Pr[R = r] \Pr[R = r | E]}{\Pr[E] \Pr[R = r | E]} \right) \\
&= (1/n) \log \left(2^h / \Pr[E] \right) \\
&= \frac{d+h}{n}
\end{aligned}$$

□

6 Proof of Main Lemmas

In this section, we shall prove Lemma 3.1 and Lemma 3.2.

These lemmas say that as long the probability of winning in the k coordinates in S is not too small, then on average, the players must be doing pretty badly on the remaining coordinates even conditioned on winning in W_S.

Without loss of generality, we assume that $S = \{n-k+1, n-k, \ldots, n\}$. We shall prove these lemmas by contradiction. Fix a strategy for $\mathcal{G}^n$. Suppose for the sake of contradiction that $\Pr[W_S]$ is high and $\mathbb{E}_{i \notin S}\left[\Pr[W_{\{i\}}|W_S]\right] > 1 - \epsilon/2$. Then we shall use the players strategy for $\mathcal{G}^n$ to get an extremely good strategy for $\mathcal{G}$, one that wins with probability more than $1-\epsilon$, and thus contradicting the bound on the value of $\mathcal{G}$.

6.1 Intuition for the proof

We first outline a natural way to use a strategy for $\mathcal{G}^n$ to get a strategy for $\mathcal{G}$, which we shall ultimately refine to complete the proof: the players decide on an index i such that given questions (X, Y) in $\mathcal{G}$, they can use shared randomness to generate $n-1$ pairs of questions such that when the questions (X, Y) are placed in the i'th coordinate, and the rest of the questions are placed in the appropriate coordinates, the resulting distribution is statistically close to the distribution $(X_1, Y_1) \ldots (X_n, Y_n)|W_S$. If the players can find such an index i, then they could just use the strategy of the players for $\mathcal{G}^n$ to win $\mathcal{G}$ in the i'th coordinate with probability more than $1-\epsilon$.

There are a couple of obstacles to getting this approach to work. The most immediately apparent obstacle is that it must be true that there exists an index i for which $(X_i, Y_i)|W_S$ is statistically close to (X, Y). This obstacle can easily be circumvented via Lemma 5.6. A more subtle issue is that the players have to generate the rest of the questions without communicating. Dealing with this issue will take up most of our effort in the proof. To understand under what circumstances it is possible for two players to generate questions that satisfy the above properties, let us first look at some simple cases. Below, let X^n denote $(X_1, \ldots, X_n)$ and Y^n denote $(Y_1, \ldots, Y_n)$.

Independent Distributions. Suppose every $(x, y) \in (\mathcal{X}, \mathcal{Y})$ was such that $\left(X^n, Y^n \middle| W_S \wedge (X_i = x)\right)$ and $\left(X^n, Y^n \middle| W_S \wedge (Y_i = y)\right)$ are both product distributions. Then given the questions (x, y), the first player can sample $\left(X^n \middle| W_S \wedge (X_i = x)\right)$ and the second player can independently sample $\left(Y^n \middle| W_S \wedge (Y_i = y)\right)$. It is then easy to see that if X, Y was statistically close to $(X_i, Y_i)|W_S$ (which we can guarantee using Lemma 5.6), the players do sample questions which are statistically close to $X^n Y^n | W_S$. Of course the assumption that we have such independence is unreasonable. In general, the first player's questions are not independent of the second players questions. Even if the game was such that (X, Y) is a product distribution, conditioning on W_S could potentially introduce complicated dependencies between the questions of the players.

Completely correlated distributions. Next suppose we could somehow prove that there exists some random variable R and functions f, g such that for every $x \in \mathcal{X}$, $y \in \mathcal{Y}$:

- Learning R would allow both players to generate the random variables they want using their inputs
$$f(R, x), g(R, y) \approx \left(X^n, Y^n \middle| W_S \wedge (X_i = x) \wedge (Y_i = y)\right)$$
- Although R may depend on (X_i, Y_i), all the information needed to generate R is contained in any one of these variables:
$$\left(R \middle| W_S \wedge (X_i = x)\right) \approx \left(R \middle| W_S \wedge (Y_i = y)\right) \approx \left(R \middle| W_S \wedge (X_i = x) \wedge (Y_i = y)\right)$$

Given these conditions, it is easy to design a protocol for the players — each player computes the distribution for R based on his question and they then use Lemma 4.1 to agree on a sample

for $\big(R\big|W_S \wedge (X_i = x) \wedge (Y_i = y)\big)$. The lemma and the second condition above guarantee that the distribution they end up with will be statistically close to the right one. Once they have generated the sample for R, they simply apply the functions f, g to generate their corresponding questions.

The solution for the general case will be a mixture of the solutions in the above two cases. We shall identify an index i and a random variable R such that:

- Fixing $R = r$ will determine at least one question of (X_i, Y_i) for every coordinate i. If A denotes the answers of one of the players in the coordinates $n-k+1, \ldots, n$, this condition guarantees that for every r, a, x, y, X^n, Y^n are independent conditioned on the event $(R, A, X_i) = (r, a, x) \wedge W_S$. Similarly X^n, Y^n are independent conditioned on the event $(R, A, Y_i) = (r, a, y) \wedge W_S$.
- Conditioned on W_S, all the information needed to generate RA given (X_i, Y_i) is contained in any one of these variables:

$$\big\{X_iY_i\big|W_S\big\}\big\{RA\big|X_i \wedge W_S\big\} \approx \big\{X_iY_iRA\big|W_S\big\} \approx \big\{X_iY_i\big|W_S\big\}\big\{RA\big|Y_i \wedge W_S\big\}$$

Once we are able to determine such R, A and prove the above properties, we shall be done. On receiving the questions x, y, the players will use the protocols from Lemma 4.1 to generate $\big(RA\big|(X_i, Y_i) = (x, y) \wedge W_S\big)$. Once they have sampled this random variable, they can generate the rest of their questions independently. This would prove that $\Pr[W_{\{i\}}|W_S]$ must be small.

The stronger results that apply to projection games in this paper come about by proving that in these kinds of games, the only way the player can win is by using a strategy that restricts itself to using a few possible answers. We can define a new sub-event of W_S that ensures that not only do the players win in the coordinates of S, they do so by using answers that have a relatively high probability. We can show that this event has an extremely high density in W_S, so that conditioning on W_S is essentially the same as conditioning on this event. Thus allows to carry out the proof as if the effective answer size of the provers is much smaller than it actually is.

6.2 The proof

Let $A = A_{n-k+1} \ldots A_n$ and $B = B_{n-k+1} \ldots B_n$ denote the answers of the players in the last k games. Let $V = V_1, V_2, \ldots, V_{n-k}$ denote uniformly random bits.

For $i = 1, 2, \ldots, n-k$, let T_i denote a random question in every coordinate:

$$T_i \stackrel{def}{=} \begin{cases} X_i & \text{if } V_i = 1, \\ Y_i & \text{if } V_i = 0. \end{cases}$$

Let U_i's denote the opposite questions:

$$U_i \stackrel{def}{=} \begin{cases} X_i & \text{if } V_i = 0, \\ Y_i & \text{if } V_i = 1. \end{cases}$$

Set $Q \stackrel{def}{=} X_{n-k+1}X_{n-k+2} \ldots X_nY_{n-k+1}Y_{n-k+2} \ldots Y_n$ — the "won" questions.

Set $R \stackrel{def}{=} VQT_1T_2 \ldots T_{n-k}$ — the "won" questions and a random question from each of the remaining question pairs.

Set $R^{-j} \stackrel{def}{=} VQT_1T_2 \ldots T_{j-1}T_{j+1} \ldots T_n$ — removing the j'th coordinate from R.

The most technical part of the proof is the following lemma:

Lemma 6.1. *Let E be any event that is determined by ABR and h, z be positive integers such that*

- $\Pr[E] \geq 2^{-\frac{\epsilon^2(n-k)}{z^2}+h}$.
- *For every r, $|\mathsf{supp}(A|R = r \wedge E)| \leq 2^h$.*

Then $\mathbb{E}_{i \notin S}\left[\Pr[W_{\{i\}}|E]\right] \leq 1 - \epsilon + 17\epsilon/z$.

Before proving this lemma, let us first see how we can use it to prove Lemma 3.1 and Lemma 3.2.

Proof of Lemma 3.1. Set $E = W_S$. It is clear that E is determined by QAB which is contained in ABR. If the game is such that each answer comes from a set of size 2^c, $|\mathsf{supp}(A|R = r \wedge E)|$ is trivially bounded by 2^{kc}. Set $h = kc$ and apply Lemma 6.1 to get the lemma. □

Next we give the proof for the case of projection games.

Proof of Lemma 3.2. Recall that in a projection game, we have the condition that there is some function f_Q, determined by the questions Q such that W_S holds exactly when $f_Q(B) = A$.

For any tuple $(a, r) \in (\mathcal{A}, \mathcal{R})$, say that (a, r) is heavy if $\Pr[A = a|R = r] \geq 2^{-h}$, where h is a parameter that we shall fix later. The intuition behind this definition is that conditioned on $R = r$, the answers $A, B|R = r$ are independent. Thus the players should be able to win the projection game with a decent probability only when they pick one of the heavy elements, and there cannot be too many of those. For instance, imagine that f was the identity function. Then it is easy to check that if A, B are independent and $\Pr[A = B]$ is γ, there must be a set of size $O(1/\gamma)$ (namely the elements with weight at least $\gamma/100$) which A lands in with high probability.

Let G denote the event that (A, R) is heavy. We shall argue that when the players win, they usually win inside the event G. Note that for every r, $A, B|R = r$ is a product distribution.

$$\Pr[W_S \wedge G^c] = \sum_{(b,r) \text{ s.t. } (f_q(b),r) \text{ is not heavy}} \Pr[R = r, B = b] \Pr[A = f_q(b)|R = r] \leq 2^{-h}$$

In particular this means that whenever W_S happens, G happens with high probability (if we pick h to be large enough):

$$\Pr[G|W_S] = \frac{\Pr[G \wedge W_S]}{\Pr[W_S]} = \frac{\Pr[W_S] - \Pr[W_S \wedge G^c]}{\Pr[W_S]} \geq 1 - 2^{-h}/\Pr[W_S]$$

Set $E = G \wedge W_S$. Again note that E is determined by ABR. For every random variable O in the space, the last inequality implies that $|O|E - O|W_S| \leq 2^{-h}/\Pr[W_S]$. In particular, for every i, $\Pr[W_{\{i\}}|W_S] \leq \Pr[W_{\{i\}}|E] + 2^{-h}/\Pr[W_S]$.

Set $h = (3/5)\frac{\epsilon^2(n-k)}{68^2}$. Then we get that:

$$\begin{aligned}
\Pr[E] &= \Pr[G|W_S] \Pr[W_S] \geq \Pr[W_S] - 2^{-h} \\
&\geq 2^{-(1/5)\frac{\epsilon^2(n-k)}{68^2}} - 2^{-3/5\frac{\epsilon^2(n-k)}{68^2}} \\
&= 2^{-(2/5)\frac{\epsilon^2(n-k)}{68^2}}(2^{(1/5)\frac{\epsilon^2(n-k)}{68^2}} - 2^{-1/5\frac{\epsilon^2(n-k)}{68^2}}) \\
&\geq 2^{-(2/5)\frac{\epsilon^2(n-k)}{68^2}} \\
&= 2^{-\frac{\epsilon^2(n-k)}{68^2}+h}
\end{aligned}$$

satisfying the first condition of Lemma 6.1 with $z = 68$.

Applying the lemma, we get that

$$\begin{aligned}
&\mathop{\mathbb{E}}_{i\notin S}\left[\Pr[W_{\{i\}}|W_S]\right] \\
&\leq \mathop{\mathbb{E}}_{i\notin S}\left[\Pr[W_{\{i\}}|E]\right] + 2^{-h}/\Pr[W_S] \\
&\leq 1-\epsilon+\epsilon/4+2^{-\frac{\epsilon^2 2(n-k)}{5(68)^2}} \\
&\leq 1-\epsilon/2
\end{aligned}$$

Where the last inequality comes from the lowerbound on $n-k$ in the hypothesis. □

Finally, we prove Lemma 6.1.

Proof of Lemma 6.1. We shall first show that in expectation over a random choice of the index i, if the players use the protocol from Lemma 4.1 to generate $AR^{-i}|E$ assuming that their questions came from the distribution $X_iY_i|E$, then with high probability they sample the same value for this variable which implies that the distribution they sample is close to

$$\{X_iY_i\}\{AR^{-i}|EX_iY_i\} \approx \{X_iY_i|E\}\{AR^{-i}|EX_iY_i\}$$

Then we shall argue that if the players complete the rest of the questions they need independently, the joint distribution of questions they get is close to $X^nY^n|E$.

We shall use the following shorthand to simplify notation: an expression like $F_i \overset{\gamma}{\approx}_{\mathbb{E}_{i\notin S}} G_i$ stands for the statement $\mathbb{E}_{i\notin S}[|F_i - G_i|] \leq \gamma$.

Claim 6.2. $\{X_iY_i|E\} \overset{\epsilon/z}{\approx}_{\mathbb{E}_{i\notin S}} \{X_iY_i\} =_{\mathbb{E}_{i\notin S}} \{XY\}$

Proof. The claim follows by Lemma 5.6 applied to the event E and the product distribution of the questions:

$$\mathop{\mathbb{E}}_{i\notin S}\left[\left|\{X_iY_i|E\} - \{X_iY_i\}\right|\right] \leq \sqrt{\frac{\epsilon^2(n-k)-h}{z^2(n-k)}} \leq \epsilon/z$$

□

We apply Corollary 5.7 to get that:

$$\mathop{\mathbb{E}}_{i\notin S}\left[\left|\{AR|E\}\{U_i|R\} - \{ARU_i|E\}\right|\right] \leq \sqrt{\frac{\epsilon^2(n-k)-h+h}{z^2(n-k)}} = \epsilon/z$$

Note that for every i, $\{AR|E\}\{U_i|R\} = \{AR|E\}\{U_i|T_iV_i\}$, since U_i is independent of all the other random variables in R. Conditioning on $V_i = 1$, by Proposition 2.4, we get

$$\{ARU_i|E\} \overset{\epsilon/z}{\approx}_{\mathbb{E}_{i\notin S}} \{AR|E\}\{U_i|T_iV_i\} \Rightarrow \{AR^{-i}Y_iX_i|E\} \overset{2\epsilon/z}{\approx}_{\mathbb{E}_{i\notin S}} \{AR^{-i}Y_i|E\}\{X_i|Y_i\} \qquad (1)$$

We can then argue that

$$\begin{aligned}
&\{X_iY_i\}\left\{AR^{-i}\middle|Y_iX_iE\right\} \\
&\overset{\epsilon/z}{\approx}_{\mathbb{E}_{i\notin S}} \left\{X_iY_i\middle|E\right\}\left\{AR^{-i}\middle|Y_iX_iE\right\} && \text{by Claim 6.2} \\
&=_{\mathbb{E}_{i\notin S}} \left\{AR^{-i}X_iY_i\middle|E\right\} && \text{rearranging} \\
&\overset{2\epsilon/z}{\approx}_{\mathbb{E}_{i\notin S}} \left\{AR^{-i}Y_i\middle|E\right\}\left\{X_i\middle|Y_i\right\} && \text{by Equation 1} \\
&=_{\mathbb{E}_{i\notin S}} \left\{Y_i\middle|E\right\}\left\{X_i\middle|Y_i\right\}\left\{AR^{-i}\middle|Y_iE\right\} && \text{rearranging} \\
&\overset{\epsilon/z}{\approx}_{\mathbb{E}_{i\notin S}} \{Y_i\}\left\{X_i\middle|Y_i\right\}\left\{AR^{-i}\middle|Y_iE\right\} && \text{by Claim 6.2} \\
&=_{\mathbb{E}_{i\notin S}} \{X_iY_i\}\left\{AR^{-i}\middle|Y_iE\right\} && \text{rearranging}
\end{aligned}$$

Repeating the argument but conditioning on $V_i = 0$, we get

Claim 6.3. $\{X_iY_i\}\left\{AR^{-i}\middle|X_iE\right\} \overset{4\epsilon/z}{\approx}_{\mathbb{E}_{i\notin S}} \{X_iY_i\}\left\{AR^{-i}\middle|X_iY_iE\right\} \overset{4\epsilon/z}{\approx}_{\mathbb{E}_{i\notin S}} \{X_iY_i\}\left\{AR^{-i}\middle|Y_iE\right\}$

At this point we have made a lot of progress. We have shown that each player has roughly the same information about the random variable AR^{-i}, even in the event E. We imagine that we run the protocol promised by Lemma 4.1 using the two players in our game, plus an additional player who gets access to both questions (x, y). All players generate AR^{-i} conditioned on E and whatever questions they have. Then by Lemma 4.1 we get a protocol which has the effect that

- player 1's variables have the distribution $\{X_i\}\left\{AR^{-i}\middle|X_iE\right\}$
- player 2's variables have the distribution $\{Y_i\}\left\{AR^{-i}\middle|Y_iE\right\}$
- player 3's variables have the distribution $\{X_iY_i\}\left\{AR^{-i}\middle|X_iY_iE\right\}$
- $\mathbb{E}_{i\notin S}\left[\Pr[\text{the players have inconsistent variables when they use the index } i]\right] \leq 2(4\epsilon/z)+2(4\epsilon/z) = 16\epsilon/z$

This means that the joint distribution that the first two players get is $16\epsilon/z$-close to the distribution of the third player. But this third player samples from a distribution that is close to the one we want:

$$\begin{aligned}
&\mathop{\mathbb{E}}_{i\notin S}\left[\left|\{X_iY_i\}\left\{AR^{-i}\middle|X_iY_iE\right\} - \left\{X_iY_iAR^{-i}\middle|E\right\}\right|\right] \\
&\leq \mathop{\mathbb{E}}_{i\notin S}\left[\left|\{X_iY_i\} - \left\{X_iY_i\middle|E\right\}\right|\right] \\
&\leq \epsilon/z && \text{by Claim 6.2}
\end{aligned}$$

These facts give that for an average i, the first two players sample from a distribution that is $17\epsilon/z$ close to the correct distribution. To end the proof, observe that for every i, x, y, a, r^{-i}, $\left(X^n, Y^n\middle|(X_i, R^{-i}, A) = (x, r^{-i}, a)\right)$ and $\left(X^n, Y^n\middle|(Y_i, R^{-i}, A) = (y, r^{-i}, a)\right)$ are both product distributions. This means that if the players sample the rest of the questions they need conditioned on the information they have, we will generate a distribution that is $17\epsilon/z$ close to $X^nY^n|E$. This gives us our final bound:

$$1 - \epsilon \geq \mathop{\mathbb{E}}_{i\notin S}\left[\Pr[W_{\{i\}}|E]\right] - 17\epsilon/z$$

□

7 The Concentration Bound

In this section we prove the following concentration bounds for parallel repetition:

Theorem 7.1 (Concentration in General Games)**.** *Let $\mathcal{G}$ be a game with value $1-\epsilon$ and answer set size 2^c. Then for any $\delta > 0$, the probability that the players can win more than a $1-\epsilon+\delta$ fraction of the games in the n-fold parallel repetition is bounded by $3\exp\left(-\frac{\delta^4 n}{162(35)^2 c}\right)$.*

Here we made no effort to optimize the constants appearing in the bound. It is conceivable that these can be improved significantly. In analogy with the the results from the earlier section, the bounds can easily be improved in the case of projection games to remove the dependence on c.

We shall rely on the following incarnation of the Chernoff bound:

Theorem 7.2 (Chernoff Bound)**.** *Let $F_1, \ldots, F_m$ be independent boolean random variables with $\Pr[F_i = 1] = \mu$. Then for every constant $\delta > 0$, $\Pr[\sum_i F_i < (1-\delta)\mu m] < \exp(\mu m \delta^2/2)$.*

Our proof will rely on the following well known facts about *supermartingales*.

Definition 7.3 (Supermartingale)**.** A sequence of real valued random variables $J_0, J_1, \ldots, J_m$ is called a *supermartingale* if for every $i > 1$, $\mathbb{E}[J_i | J_0, \ldots, J_{i-1}] \leq J_{i-1}$.

We have the following concentration bound for super-martingales:

Theorem 7.4 (Azuma-Hoeffding Inequality)**.** *If $J_0, \ldots, J_m$ is a supermartingale with $J_{i+1} - J_i \leq 1$,*

$$\Pr[J_m > J_0 + \alpha] \leq \exp(-\alpha^2 m/2)$$

Armed with this inequality, we can now prove our concentration bound. Recall that Lemma 6.1 was the heart of the proofs of the parallel repetition theorems that we have seen. The idea behind the concentration bound is to exploit the fact that this lemma shows that for every set S, not only is there a single index i for which $\Pr[W_{\{i\}}|W_S]$ is small, but that even if we pick a random i, the expected value of this probability is small.

Imagine that the referee in the game doesn't check that the players win in all the coordinates, but merely samples a few coordinates at random and checks that they win in those coordinates. Then the random variables of whether or not the players have won should behave somewhat like a martingale — conditioned on the outcome in a few coordinates, the outcome in the next coordinate should still be biased towards losing. We can use the Azuma-Hoeffding inequality to bound the probability that the referee sees an unusually large fraction of won games in her random sample. On the other hand, by the Chernoff bound the fraction of won games that the referee sees is a very good estimate for the total fraction of games that the players won. We can use these two facts to show that the players win a larger fraction of games with only negligible probability.

Proof of Theorem 7.1. Let δ be as in the hypothesis. Set $\nu \stackrel{def}{=} \delta/3$, $z \stackrel{def}{=} 34\epsilon/\nu$ and $m \stackrel{def}{=} \frac{\nu^2 n}{35^2 c}$.

Let $V \stackrel{def}{=} V_1, \ldots, V_m$ be independent random integers from $[n]$ and let V^i denote this sequence truncated after i integers. For any integer i, let W_i denote the random variable that takes the value 1 if the event $W_{\{i\}}$ holds and 0 otherwise. Set $D \stackrel{def}{=} W_{V_1}, \ldots, W_{V_M}$ and let D^i be this sequence truncated after the i'th bit.

Given a vector $d^i = (d_1, \ldots, d_i) \in \{0,1\}^i$ and $v^i = (v_1, \ldots, v_i) \in [n]^i$, say that (d^i, v^i) are *typical* if

$$\Pr\left[D^i = d^i | V^i = v^i\right] \geq 2^{-2m}$$

Let G_i denote the event that (D^i, V^i) are typical. Then note that $G_i^c \Rightarrow G_{i+1}^c$ and

$$\Pr[G_m^c] = \sum_{\text{atypical } (v,d)} \Pr[D = d | V = v] \Pr[V = v] < 2^{-m}$$

Set J_0 to be the random variable taking the value 0 with probability 1. Set $\gamma \stackrel{def}{=} \epsilon - 17\epsilon/z - m/n \geq \epsilon - \nu/2 - \nu/2 = \epsilon - \nu$ by our choice of parameters. For every $i = 1, \ldots, m$, set

$$J_i \stackrel{def}{=} \begin{cases} J_{i-1} + \gamma & \text{if } W_{\{V_i\}} \wedge G_{i-1}, \\ J_{i-1} + \gamma - 1 & \text{otherwise.} \end{cases}$$

Given any fixing (d^{i-1}, v^{i-1}) of (D^{i-1}, V^{i-1}), if (d^{i-1}, v^{i-1}) are not typical, then $J_i = J_{i-1} + \gamma - 1 < J_{i-1}$. On the other hand, if (d^{i-1}, v^{i-1}) are typical, let E denote the event $V^{i-1} = v^{i-1} \wedge D^{i-1} = d^{i-1}$. Then $\Pr[E] \geq 2^{-2m} \geq 2^{-\frac{\epsilon^2(n-m)}{z^2} + c}$ by our choice of parameters. Also, E is determined by the answers and questions in the games corresponding to $V_1, \ldots, V_{i-1}$. Lemma 6.1 then gives: $\mathbb{E}\left[\left(D_i \middle| V^{i-1} = v^{i-1} \wedge D^{i-1} = d^{i-1}\right)\right] \leq m/n + 1 - \epsilon + 17\epsilon/z = 1 - \gamma$, which implies that $\mathbb{E}\left[J_i | G_{i-1}, J_0, \ldots, J_{i-1}\right] \leq J_{i-1}$. Thus,

Claim 7.5. *$J_0, \ldots, J_m$ is a supermartingale.*

Theorem 7.4 then gives us the bound:

$$\begin{aligned}
&\Pr\left[\sum_{i=1}^m D_i \geq (1 - \epsilon + 2\nu)m\right] \\
&\leq \Pr[G_m^c] + \Pr[J_m \geq (1 - \epsilon + 2\nu)m\gamma - (\epsilon - 2\nu)(1 - \gamma)m] \\
&\leq 2^{-m} + \Pr[J_m \geq m(\gamma - (\epsilon - 2\nu))] \\
&\leq 2^{-m} + \Pr[J_m \geq m(\epsilon - \nu - (\epsilon - 2\nu))] \\
&\leq 2^{-m} + \exp(-\nu^2 m/2) \\
&< 2\exp(-\nu^2 m/2)
\end{aligned}$$

To finish the proof, we note that by the Chernoff bound, we expect that the vector D gives a good estimate for the fraction of games that were won by the players. Specifically, Theorem 7.2 promises that:

$$\begin{aligned}
&\Pr\left[\sum_{i=1}^m D_i < (1 - \epsilon + 2\nu)m \wedge \sum_{i=1}^n W_i \geq (1 - \epsilon + 3\nu)n\right] \\
&\leq \Pr\left[\left(\sum_{i=1}^m W_{V_i} < (1 - \epsilon + 2\nu)m \middle| \sum_{i=1}^n W_i \geq (1 - \epsilon + 3\nu)n\right)\right] \\
&< \exp\left(-m\frac{1 - \epsilon + 3\nu}{2}\left(1 - \frac{1 - \epsilon + 2\nu}{1 - \epsilon + 3\nu}\right)^2\right) \\
&= \exp\left(-m\frac{1 - \epsilon + 3\nu}{2}\left(\frac{\nu}{1 - \epsilon + 3\nu}\right)^2\right) \\
&< \exp(-m\nu^2/2)
\end{aligned}$$

This gives us our final estimate:

$$\begin{aligned}
&\Pr\left[\sum_{i=1}^{n} W_i \geq (1-\epsilon+\delta)n\right] \\
&= \Pr\left[\sum_{i=1}^{n} W_i \geq (1-\epsilon+3\nu)n\right] \\
&\leq \Pr\left[\left(\sum_{i=1}^{m} D_i < (1-\epsilon+2\nu)m\right) \wedge \left(\sum_{i=1}^{n} W_i \geq (1-\epsilon+3\nu)n\right)\right] + \Pr\left[\sum_{i=1}^{m} D_i \geq (1-\epsilon+2\nu)m\right] \\
&\leq 3\exp(-m\nu^2/2) \\
&= 3\exp\left(-\frac{\delta^4 n}{162(35)^2 c}\right)
\end{aligned}$$

■

8 Acknowledgments

Thanks to Boaz Barak, Guy Kindler, Venkatesan Guruswami, Scott Aaronson, Avi Wigderson, David Zuckerman and Parikshit Gopalan for useful discussions. Boaz Barak gave the intuition for the proof of Holenstein's lemma sketched in Lemma 4.1. Venkatesan Guruswami pointed out the application of our work to the unique games conjecture. Scott Aaronson told us the motivation for getting a concentration bound. We would also like to thank the anonymous referee who suggested how to remove a required lowerbound on n from our main theorem in an earlier version of this work.

References

[ALM+98] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM*, 45, 1998.

[AS98] Sanjeev Arora and Shmuel Safra. Probabilistic checking of proofs: A new characterization of NP. *Journal of the ACM*, 45(1):70–122, January 1998.

[Aus07] Per Austrin. Balanced max 2-sat might not be the hardest. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, pages 189–197. ACM, 2007.

[BCH+02] Jonathan Barrett, Daniel Collins, Lucien Hardy, Adrian Kent, and Sandu Popescu. Quantum nonlocality, bell inequalities and the memory loophole. *Physical Review A*, 66:042111, 2002.

[Bel64] John S. Bell. On the Einstein-Podolsky-Rosen paradox. *Physics*, 1(3):195–290, 1964.

[BGKW88] Michael Ben-Or, Shafi Goldwasser, Joe Kilian, and Avi Wigderson. Multi-prover interactive proofs: How to remove intractability assumptions. In *Proceedings of the 20th Annual ACM Symposium on Theory of Computing*, pages 113–131, 1988.

[BGKW89] Michael Ben-Or, Shafi Goldwasser, Joe Kilian, and Avi Wigderson. Efficient identification schemes using two prover interactive proofs. In *Advances in Cryptology — CRYPTO '92, 12th Annual International Cryptology Conference, Proceedings*, 1989.

[CMM06] Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Near-optimal algorithms for unique games. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing*, 2006.

[CKK+06] Shuchi Chawla, Robert Krauthgamer, Ravi Kumar, Yuval Rabani, and D. Sivakumar. On the hardness of approximating multicut and sparsest-cut. *Proceedings of the 21th Annual IEEE Conference on Computational Complexity*, 15, 2006.

[CMM06] Eden Chlamtac, Konstantin Makarychev, and Yury Makarychev. How to play unique games using embeddings. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science*, 2006.

[CHSH69] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.*, 23(15):880–884, Oct 1969.

[DMR06] Irit Dinur, Elchanan Mossel, and Oded Regev. Conditional hardness for approximate coloring. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing*. ACM, 2006.

[DFK+92] Cynthia Dwork, Uriel Feige, Joe Kilian, Moni Naor, and Shmuel Safra. Low communication 2-prover zero-knowledge proofs for NP. In *Advances in Cryptology — CRYPTO '92, 12th Annual International Cryptology Conference, Proceedings*, 1992.

[FGL+91] Uriel Feige, Shafi Goldwasser, Laszlo Lovasz, Shmuel Safra, and Mario Szegedy. Approximating clique is almost NP-complete (preliminary version). In *Proceedings of the 32nd Annual IEEE Symposium on Foundations of Computer Science*, 1991.

[FL92] Uriel Feige and Laszlo Lovasz. Two-prover one-round proof systems: their power and their problems. In *Proceedings of the 24th Annual ACM Symposium on Theory of Computing*, pages 733–744, 1992.

[FV02] Uriel Feige and Oleg Verbitsky. Error reduction by parallel repetition–A negative result. *Combinatorica*, 22, 2002.

[Gil03] Richard D. Gill. Accardi contra bell (cum mundi): The impossible coupling. *IMS LECTURE NOTES-MONOGRAPH SERIES*, 42, 2003.

[GW95] Michel Goemans and David Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM*, 42, 1995.

[GT06] Anupam Gupta and Kunal Talwar. Approximating unique games. In *SODA*, pages 99–106. ACM Press, 2006.

[Hol07] Thomas Holenstein. Parallel repetition: Simplifications and the no-signaling case. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, 2007.

[Kho02] Subhash Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, pages 767–775, 2002.

[KKMO04] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell. Optimal inapproximability results for max-cut and other 2-variable CSPs? In *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*, 2004.

[KR03] Subhash Khot and Oded Regev. Vertex cover might be hard to approximate to within 2-ϵ. In *IEEE Conference on Computational Complexity*, page 379. IEEE Computer Society, 2003.

[KV05] Subhash Khot and Nisheeth Vishnoi. The unique games conjecture, integrality gap for cut problems and embeddability of negative type metrics into ℓ_1. In *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, 2005.

[LS95] Dror Lapidot and Adi Shamir. A one-round, two-prover, zero-knowledge protocol for NP. *Combinatorica*, 15, 1995.

[LY93] Carsten Lund and Mihalis Yannakakis. On the hardness of approximating minimization problems. In *Proceedings of the 25th Annual ACM Symposium on Theory of Computing*, pages 286–293, 1993.

[MOO05] Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz. Noise stability of functions with low in.uences invariance and optimality. In *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, pages 21–30. IEEE Computer Society, 2005.

[PRW97] Itzhak Parnafes, Ran Raz, and Avi Wigderson. Direct product results and the GCD problem, in old and new communication models. In *Proceedings of the 29th Annual ACM Symposium on Theory of Computing*, pages 363–372, 1997.

[Raz98] Ran Raz. A parallel repetition theorem. *SIAM Journal on Computing*, 27(3):763–803, June 1998.

[Raz08] Ran Raz. A counterexample to strong parallel repetition. *Manuscript*, 2008.

[Tre05] Luca Trevisan. Approximation algorithms for unique games. In *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, 2005.

[Ver94] Oleg Verbitsky. Towards the parallel repetition conjecture. In *Structure in Complexity Theory Conference*, pages 304–307, 1994.