The Parallel Repetition Theorem and Related Results

Rakesh Venkat

Project Report
Under the supervision of Prof. Jaikumar Radhakrishnan

Tata Institute of Fundamental Research, Mumbai

Abstract

In a 2-Prover 1-Round Game, a verifier draws a pair of questions $(X, Y)$ from a distribution $D$ and sends one each to two co-operating, non-communicating players who need to respond back with answers $A, B$. The verifier checks the answers using a known predicate $V(X, Y, A, B)$, and declares a win or loss. The aim of the players is to plan a strategy to win the game with the highest probability over the set of questions. The $n$-fold parallel repetition of the game has the verifier drawing $n$ pairs of questions $(\bar{X}, \bar{Y}) \sim D^n$, i.i.d., and sending each player an $n$-tuple of questions. The players have to respond with $n$ answers each, one for each co-ordinate. The Parallel Repetition theorem proven by Raz states that the probability of winning the repeated game in all co-ordinates drops exponentially with $n$. The theorem was a key result used in proving various hardness of approximation results. In this report, we look to survey results closely related to this theorem, and the implications it has in various areas of computational complexity.

Contents

1 Introduction .................................................. 2

2 Motivation And Relation To Hardness Of Approximation .......................... 4
   2.1 2-Prover 1-Round games .................................. 4
   2.2 PCPs and Hardness of Approximation ..................... 8
   2.3 Feige and Verbitsky’s Results .............................. 11
   2.4 Feige and Kilian’s Repetition theorem .................... 13

3 Proof of the Parallel Repetition Theorem ..................................... 16
   3.1 The Theorem and Proof Methodology ...................... 16
   3.2 The Proof .................................................. 16
   3.3 Projection Games .......................................... 21
   3.4 Concentration Result ...................................... 23

4 Raz’s Counterexample and Unique Games ....................................... 27
   4.1 Counterexample to Strong Parallel Repetition .......... 27
   4.2 QP and SDP formulations ................................. 31
   4.3 The Unique Games conjecture ............................ 34

5 Related Results ................................................. 37
   5.1 Tiling in High Dimensions ................................. 37
   5.2 Other Results ............................................... 41
1 Introduction

A common question in computational complexity is the following: Given a computational model, and a function \( f : \mathcal{D} \to \mathcal{R} \) that can be computed in this model (possibly with some error probability \( \gamma \) over \( \mathcal{D} \)), how “difficult” is it to compute independent copies of this function? That is, given inputs \( x_1, x_2, \ldots, x_k \) drawn independently from the input domain \( \mathcal{D} \), how do we compute \( F_k = (f(x_1), \ldots, f(x_k)) \) in a way “better” than computing each copy independently? We call the function \( F_k \) the direct product of the function \( f \), taken \( k \) times.

There are two types of questions that one can normally ask in such a setting. One is to obtain bounds on the resources required, under the complexity measure of the model, of computing the direct product function. The other is to keep a bound on the resources available, and try to maximize the quality of computation; that is, maximize the probability over the input space, of computing the direct product function correctly. In our setting, we shall be concerned with the latter question. The naive way of computing \( F_k \) (computing each instance independently) requires resources \( k \) times the resource of computing \( f \). The success probability, in this case, decreases as \((1-\gamma)^k\). The questions posed essentially ask if this is the best one can do.

Results in these directions are usually used to amplify the error probability associated with the computation of a certain problem. For instance, in the area of hardness amplification, Yao’s XOR lemma informally states that if a function \( f : \{0,1\}^n \to \{0,1\} \) cannot be computed with probability greater than \( 1-\delta \) by small circuits, then the new function \( F = \bigoplus_{i \in [k]} f(x_i) \) where each \( x_i \) is chosen uniformly and independently from \( \{0,1\}^n \) cannot be computed by small circuits with probability better than \( \frac{1}{2} + 2(1-\delta)^k \).

Similar direct product results appear in other areas, such as communication complexity (e.g. the success probability of computing the direct product function \( f^{\otimes k} \) using a collection of protocols of depth at most \( c \), one for each co-ordinate, is \( 2^{-\Omega(k/c)} \) if \( c \) is the communication complexity of computing \( f \) up to constant error over the inputs [PRW97]).

The setting that we will be considering here is a 2-Prover 1-Round game. The motivation for considering such a setting comes from the characterization of the complexity class NP using Interactive Proofs, made in a long line of work culminating in the PCP Theorem [AS98, ALM+98]. Consider the MAX-3SAT problem: maximizing the number of satisfied clauses to a 3CNF formula using some assignment to the variables. The PCP Theorem itself only gives a statement of the following form regarding this problem:

**Corollary.** (Of the PCP Theorem, Informal) There is a constant \( 1 > \epsilon_0 > 0 \) such that given a 3CNF formula \( \psi \), it is NP-hard to distinguish between the cases where \( \psi \) is satisfiable, and where any assignment satisfies less than \( \epsilon_0 \)-fraction of the clauses. That is, unless \( P = NP \), there is no polynomial time \( \epsilon_0 \)-approximation algorithm for finding an assignment that maximizes the number of satisfied clauses in a 3CNF formula.

Ideally, we would like the constant \( \epsilon_0 \) to be pushed down as far as possible, since it would imply stronger hardness of approximation results for MAX-3SAT, and through reductions, hardness results for other problems. To boost the gap in the above statement from \((\epsilon_0,1)\) to \((\epsilon,1)\) for as small an \( \epsilon \) as possible, one employs a direct product construction combined with a parallel repetition theorem for that construction (e.g. [Raz98, FK00]). While we defer the statement and the precise setting to the following chapter, we can summarize the final effect as follows: by using a direct-product-like setting, we can come up with a polynomial-time reduction \( R \), that using \( \psi \) produces 3CNF instances \( R(\psi) \) such that:

- \( \psi \) is satisfiable \( \implies R(\psi) \) is satisfiable
- \( OPT(\psi) \leq \epsilon_0 \implies OPT(R(\psi)) \leq \epsilon \)

The running time of the reduction, and the size of the produced instance depend upon \( \epsilon \): roughly speaking, the number of copies \( k \) that one needs to use in the direct-product, and the resulting blow-up
in the size of the domain from $\mathcal{D}$ to $\mathcal{D}^k$ and the range from $\mathcal{R}$ to $\mathcal{R}^k$ needs to be carefully monitored. The better the parameters in the parallel repetition theorem, the better results we get in terms of hardness of approximation. In particular, the rate of decay of the success probability with $k$ and $\epsilon$ is of special interest.

This report focuses on a parallel repetition theorem first proven by Raz\cite{Raz98}, and its use in various problems in computational complexity theory. In Section 2, we describe the setting, giving the background of how a parallel repetition theorem implies strong hardness of approximation results. In Section 3, we discuss the proof of the parallel repetition theorem, and related results for the special case of projection games and concentration bounds in this setting. In Section 4, we look at an example of a specific game where bounds are tight, and see how the result relates to the Unique Games Conjecture. Finally, in Section 5, we look at related results, notably, a surprising connection to the geometric problem of tiling in high dimensions.
# 2 Motivation And Relation To Hardness Of Approximation

In this section, we make the setting and the description given in the introduction precise, and state a closely related result that was proven earlier by Feige and Kilian. We start with the setting that the theorem will be used in: a 2-Prover 1-Round game.

## 2.1 2-Prover 1-Round games

We first formalize the notion of 2-Prover 1-Round games. There are three parties in such a setting: two players or provers (usually named P1 and P2) who co-operate with each other, to win against a verifier (usually named V) in a game. The game $G$ consists of the following parameters:

- A set of questions $\mathcal{X}$ for $P_1$, a set of questions $\mathcal{Y}$ for $P_2$; and a joint distribution $D$ (or $P_{XY}$) over the set of question pairs $\mathcal{X} \times \mathcal{Y}$
- A set of answers $A$ for $P_1$, a set of answers $B$ for $P_2$
- A boolean predicate $V : \mathcal{X} \times \mathcal{Y} \times A \times B \rightarrow \{0, 1\}$

All the sets mentioned above are finite. The distribution $D$ and the predicate $V$ are known to both the players beforehand, however, during the game, the provers are not allowed to communicate with each other. The game proceeds as follows: the verifier draws a pair of questions $(x, y) \sim D$ and sends $x$ to $P_1$ and $y$ to $P_2$. $P_1$ has to come up with an answer $a$ to $x$; and $P_2$ has to come up with an answer $b$ to $y$ that they send back to the verifier. The verifier then checks the predicate $V(x, y, a, b)$ and declares a win or a loss depending on the predicate returning 1 or 0 respectively.

Since the players are not allowed to communicate with each other during the game itself (and hence cannot know the other player’s question), they have to come up with a strategy to win the game. A strategy is a pair of functions $f_1 : \mathcal{X} \rightarrow A$ and $f_2 : \mathcal{Y} \rightarrow B$, that $P_1$ and $P_2$ respectively use to answer their questions.

**Definition 2.1. (Value of a game)** The value of a game $G$ is the highest probability, over $D$ with which the players can win the game. That is,

$$\text{val}(G) = \max_{f_1, f_2} \Pr_{(x, y) \sim D} [V(x, y, f_1(x), f_2(y)) == 1]$$

$$= \max_{f_1, f_2} \mathbb{E}_D[V(x, y, f_1(x), f_2(y)) == 1]$$

We will call a game $G$ non-trivial if $\text{val}(G) \in (0, 1)$.

**Role of Randomness**

The provers may be allowed to use randomized strategies to answer their questions. There are two kinds of randomness one might consider: private randomness, where each player has his/her own supply of coins to toss, or public randomness where they both have access to a shared source $R$ of random bits. With a little thought, it can be reasoned that public randomness subsumes private randomness, since each player could simply ignore the random bits intended for the other player while answering the received question. It is important to observe that such shared randomness, if used, should be initialized before the game begins, and must not be a function of the questions received, simply because players would not know each other’s questions.

To accommodate randomness, the players now have strategies that run as $f_1 : \mathcal{X} \times R \rightarrow A$ and $f_2 : \mathcal{Y} \times R \rightarrow B$. The probability of winning the game $G$ will now be over the randomness in $R$ too.

However, randomized strategies do not really help the players, since they are simply a convex combination of deterministic strategies. So, there is a particular setting of the random source that yields a strategy that is at least as good as the average.
\[ \text{val}(G) = \max_{f_1, f_2} \Pr_{(X,Y) \sim \mathcal{D}, R \sim \mathcal{R}} [V(X, Y, f_1(X, R), f_2(Y, R) = 1] \]
\[ = \max_{f_1, f_2} \mathbb{E}_{\mathcal{D}, R} [V(X, Y, f_1(X, R), f_2(Y, R))] \]
\[ = \max_{f_1, f_2} \mathbb{E}_{D} \mathbb{E}_{R} [V(X, Y, f_1(X, R), f_2(Y, R))] \]
\[ \leq \max_{f_1, f_2} \mathbb{E}_{D} [V(X, Y, f_1(X, r), f_2(Y, r))] \text{ for some } r \in \mathcal{R} \]

### 2.1.1 Examples of formulation of problems as 2-Prover 1-Round games

Consider the problem of MAX-3SAT: Given a 3CNF formula \( \psi \), find out the maximum fraction of satisfiable clauses by any given assignment to the variables. Given a formula \( \psi \), denote \( \text{sat}(\psi) \) to be the fraction of satisfiable clauses in \( \psi \). We show how to give a 2-Prover 1-Round game instance, whose value is closely related to that of \( \psi \). Let \( \psi \) be over \( n \) variables \( x_1, \ldots, x_n \) and have \( m \) clauses \( C_1 \ldots C_m \). Consider the following game \( G_{\psi} \):

- **P1** receives a variable \( X \) from \( x_1, \ldots, x_n \), and has to answer back with an assignment for this variable.
- **P2** receives a clause \( Y \) from \( C_1, \ldots, C_m \), and has to answer back with an assignment to the three variables in this clause.
- The Verifier chooses questions as follows: Pick a clause uniformly at random, and send it to **P2**. Choose one of the variables in the clause uniformly at random and send it to **P1**.
- The Verifier checks if: (a) **P2**’s assignment satisfies \( Y \), and (b) The players assignments are consistent on \( X \).

It is easy to see that if the provers play according to the assignment that satisfies the maximum number of clauses in \( \psi \), then they win with probability \( \text{sat}(\psi) \). However, the provers might cheat, i.e. be inconsistent on their variable assignments. Suppose **P2** is inconsistent with **P1** on an \( \alpha \) fraction of the clauses (on at least one variable). Assume that **P2** always answers back with a satisfying assignment for his clause, this is obviously the worst case. **P2** gets caught when the dubious clause is selected (this happens with probability \( \alpha \)), and further when a variable on which he cheated with **P2** is chosen, which happens with probability at least 1/3. Thus, they fail with probability \( \alpha \times \frac{1}{3} \). Thus, the success probability is \( \leq 1 - \alpha/3 \). However, \( \alpha \) is at most \( 1 - \text{sat}(\psi) \). Hence, we have \( \text{val}(G_{\psi}) \leq \frac{2}{3} + \frac{\text{sat}(\psi)}{3} \).

### MAX-CUT as a 2-Prover 1-Round game

MAX-CUT is the (NP-complete) problem of finding the size of the maximum cut on a graph \( G(V, E) \). Let \( \text{max} - \text{cut}(G) \) denote the value of the maximum cut on a given graph \( G \) (as a fraction of \( |E| \)). Consider the following 2-Prover 1-Round game \( G_{G} \) played on the graph \( G \) with the vertex set numbered as \([n]\):

- **P1** receives a vertex \( v \), and he has to answer back on which side of the cut \((0/1)\) it is on (call this answer \( a \)).
- **P2** receives an edge \( \{v_1, v_2\} \), and he has to answer back as to which side of the cut \((0/1)\) the smaller vertex on the edge is on (call this answer \( b \)).
- The Verifier draws the edge uniformly at random from \( E \), sends it to **P2**, draws a vertex from \( E \) uniformly at random, sends it to **P1**.
- The Verifier checks if \( v = \min \{v_1, v_2\} \) then is \( a = b \). Else, if \( v = \max \{v_1, v_2\} \) then is \( a \neq b \).
Again, if the provers play with a consistent strategy, then they win with probability at most $\max - \text{cut}(G)$. Suppose not, and if $P_2$ is inconsistent with $P_1$ on an $\alpha$ fraction of the edges. Then the players lose if the verifier chooses an inconsistent edge (happens with probability $\alpha$), and further, asks $P_1$ a bad vertex on this edge (happens with probability $\geq 1/2$). Thus, they lose with probability at least $\alpha/2$. Since $\alpha$ is at least $\max - \text{cut}(G)$, we have that $\Pr(\text{error}) \geq (\max - \text{cut}(G))/2$. Analogous to the MAX-3SAT case, there is a simple strategy that wins with this probability: fix a maximum cut in $G$, and on edges that do not go across the cut, $P_2$ flips the bit on the smaller vertex for his answers. Thus,

$$\text{val}(G_G) = \frac{1}{2} + \frac{\max - \text{cut}(G)}{2}$$

**Remark 2.2.** We note that for any given set of questions $(v, e)$ drawn by the verifier, the verification predicate is a permutation on the answer sets of the two players. Such a game is a unique game, and we will have more to say about Unique Games, and MAX-CUT in particular in Section 4.3.

### 2.1.2 The Parallel Repetition of the game

The parallel repetition of the game $G$, taken $n$ times, is also a 2-Prover 1-Round game, denoted by $G^\otimes n$. Instead of asking a single question to each player, the verifier now draws $n$ pairs of questions from the distribution $D$ in an i.i.d. manner, and sends each player an $n$-tuple of questions. They are supposed to answer all the questions correctly. Formally, the parameters of $G^\otimes n$ are:

- $P_1$ receives a question $x = (x_1, \ldots, x_n) \in X^n$, $P_2$ receives $y = (y_1, \ldots, y_n) \in Y^n$. The distribution over the question set is $D^n$.
- $P_1$ has to answer back with $a = (a_1, \ldots, a_n) \in A^n$, $P_2$ with $b = (b_1, \ldots, b_n) \in B^n$.
- The verification predicate is $V_n(x, y, a, b) = \bigwedge_{i \in [n]} V(x_i, y_i, a_i, b_i)$

To win $G^\otimes n$, the players could simply use the strategy for $G$ on each of the $n$ co-ordinates of the repeated game independently. This yields:

$$\text{val}(G^\otimes n) \geq (\text{val}(G))^n$$

However, the players are allowed to actually correlate their answers across co-ordinates. This could possibly lead to a better strategy, as demonstrated by the following counterexample due to Feige and Lovasz.

**Example 2.3.** Counter-example [FL92] : Consider a game $G$ that proceeds as follows:
Feige and Kilian then proved a theorem that shows that the rate does drop inverse-polynomially in any one of them has received

- Alice could claim “Bob has received coin 0”, which is answer (B, 0). Or she could say “Alice has received coin 1” (A, 1)
- The answer set of the players is thus in $A \times \{0, 1\}$

The verifier’s predicate on receiving answers $a, b$ is:

$\cdot \ a = b = (\alpha, i)$ for some $(\alpha, i) \in A$ and Player $\alpha$ has actually received input $i$

Claim. $\text{val}(G) = \frac{1}{2}$

Proof. On any question they receive, the provers have to decide whom to make a claim about. If they make a claim about the same prover $\alpha$, they can succeed with probability at most $\frac{1}{2}$, since the other prover knows nothing about $\alpha$’s coin. Thus, they win overall with probability at most $\frac{1}{2}$. The obvious way to achieve this value would be to decide beforehand whom to claim about (say $B$), and Alice, ignoring her input, always answers $(B, 0)$ while Bob answers with $(B, \text{coin he has received})$.

Claim. $\text{val}(G^{\otimes 2}) \geq \frac{1}{2}$

Proof. Suppose Alice received $(i, i')$ and Bob received $(j, j')$. Alice answers back with $(A, i), (B, i)$. Bob answers back with $(A, j'), (B, j')$. This is correct whenever $i = j'$, which happens with probability $\frac{1}{2}$.

One can observe with a little work (using induction), that the value of the above example game repeated $n$ times is still $\text{val}(G^{\otimes n}) = (\frac{1}{2})^{\lceil n/2 \rceil}$. (The strategy is simply to use the above 2 coordinate strategy independently on blocks of two). This raises the question: Is it possible to actually show an exponential decay in the value of a non-trivial 2-Prover 1-Round game under parallel repetition? In fact, showing a decay of any sort was itself a challenge. In an early result, Verbitsky showed the following:

**Theorem 2.5.** (Parallel Repetition, Raz [Raz98]) There exists a global function $W : [0, 1] \rightarrow [0, 1]$, with $z < 1 \implies W(z) < 1$ such that for any game $G$, with answer set size $s \geq 2$, we have

$$\text{val}(G^{\otimes n}) \leq W(\text{val}(G))^{n/\log_2 s}$$

Here, $W(z) \rightarrow 0$ as $z \rightarrow 0$ and $W(z) \rightarrow 1$ as $z \rightarrow 1$.

Although the form of $W$ was not explicitly mentioned in the original paper, it has since been deduced to satisfy the following: if $\text{val}(G) = (1 - \epsilon)$, then

$$W(\text{val}(G)) \leq (1 - \epsilon^{32})^{\text{const}}$$

$$\implies \text{val}(G^{\otimes n}) \leq (1 - \epsilon^{32})^{\text{const}. n/\log |s|}$$

Looking at the theorem, one might ask questions of the following kind: Is the exponent of 32 the best possible one? More mysteriously, why does the answer set size play a role in the repeated game? In what follows, we will try to motivate why the above questions are of interest, and look at some results that try to answer them. But before that, in the next section, we discuss why the 2-Prover 1-Round game is central to the notion of hardness of approximation of problems.
2.2 PCPs and Hardness of Approximation

We now briefly attempt to explain how and why the interest in formulating a parallel repetition theorem arose. In order to do this, we will need to define what are called Probabilistically Checkable Proofs (PCPs).

**Definition 2.6. (PCP System)** Let $L \subseteq \{0,1\}^*$ be a language and $r,q : \mathbb{N} \rightarrow \mathbb{N}$ be functions, and $c,s \in [0,1]$. $L$ is said to have a $(r(n),q(n))$-restricted verifier with completeness $c$ and soundness $s$, if there is a probabilistic polynomial time algorithm $V$ (verifier) that when given some $x \in \{0,1\}^*$ with $|x| = n$ as input, and random access to a string (which we call the proof) $\pi \in \{0,1\}^*$ of length at most $q(n)2^{r(n)}$, does the following:

- Tosses $r(n)$ coins, and decides on $q(n)$ locations in the proof to query
- Examines $\pi$ on these locations, and gives a verdict of 1 (accept) or 0 (reject). We will denote the output on proof $\pi$, input $x$ and randomness $R$ by $V^\pi[x;R]$

The output of the verifier must satisfy the following:

- **(Completeness)** If $x \in L$ then $\exists \pi : Pr_{R}[V^\pi[x;R] = 1] \geq c$
- **(Soundness)** If $x \notin L$ then $\forall \pi : Pr_{R}[V^\pi[x;R] = 1] \leq s$

We say $L$ is in $PCP_{c,s}[r(n),q(n)]$, if $L$ has a $(r(n),q(n))$-restricted verifier with completeness $c$ and soundness $s$. If $c,s$ are not explicitly mentioned, it is assumed that $c=1,s=\frac{1}{2}$.

Trivially, $NP = PCP[0\text{, poly}(n)] = PCP[\log n, \text{poly}(n)]$. It turns out that characterizing the class NP through PCPs with better parameters leads to non-trivial hardness of approximation results. We now turn to looking at this connection in more detail.

2.2.1 PCPs and Hardness of Approximation

The seminal work of Feige et. al. [FGL+96] gave the first connection between interactive proofs and hardness of approximation, by showing how the existence of Interactive Proofs for certain languages implied the hardness of approximating the size of the maximum clique in a given graph. Following this, a series of works looked at optimizing the PCP parameters in the characterization of NP, which led to the PCP Theorem:

**Theorem 2.7. (PCP Theorem)** [AS98, ALM+98] $NP = PCP[O(\log n), O(1)]$.

More precisely, $\exists Q$ such that $\forall L \in NP \exists cL$ such that $L \in PCP[cL \log n, Q]$

Consider the following decision problem, which we call $gap_\alpha - MAX3SAT$

**Definition 2.8. ($gap_\alpha - MAX3SAT$)** Let $\phi$ be a 3-CNF formula with $m$ clauses. The decision problem is to distinguish between the following cases:

- **YES** = $\{\phi : \phi$ is satisfiable$\}$
- **NO** = $\{\phi :$ any assignment to variables satisfies $\leq \alpha m$ clauses$\}$

(We don’t care about instances that belong to neither of these classes).

**Remark 2.9.** For the case of MAX-3SAT, it is easy to find an $\frac{2}{7}$-approximation to the solution : Setting every variable randomly to 0 or 1 satisfies $\frac{2}{7}$ fraction of the clauses in expectation. By using the method of conditional expectations, one can derandomize the above procedure and find an assignment that satisfies as many clauses. Hence, $gap_\alpha - MAX3SAT$ is easy for $\alpha \leq \frac{2}{7}$.

It turns out that we can restate the PCP theorem in the following sense:
Theorem 2.10. (PCP Theorem, Hardness view) \( \exists \alpha_0 \in (0, 1) \) such that \( \text{gap}_{\alpha_0} - \text{MAX3SAT} \) is NP-hard.

The above statement implies that it is NP-hard to find an assignment to a 3SAT instance that satisfies more than \( \alpha_0 \) fraction of the clauses. We do not give a proof of equivalence of Theorem 2.7 and Theorem 2.10 here.

In order to show NP-hardness results for approximating other problems, one would ideally have to show a reduction from SAT instances. For instance, consider the decision problem of approximating the size of the maximum clique in a graph \( G \):

**Definition 2.11.** \((\text{gap}_\rho - \text{MAXCLIQUE})\) Given a graph \( G \) and an integer \( k \), decide which of the following two classes it belongs to:

\[
\begin{align*}
\text{YES} & = \{ G | G \text{ has a clique of size } k \} \\
\text{NO} & = \{ G | G \text{ has no clique of size } \geq pk \}
\end{align*}
\]

To show NP-hardness for this for some given \( \rho \), we would have to give a polytime reduction \( f \) that given a 3CNF formula \( \phi \), would produce a graph \( G \) and an integer \( k \) satisfying

\[
\begin{align*}
\phi \in 3\text{SAT} & \implies f(\phi) \equiv (G, k) : G \text{ has a clique of size } \geq k \\
\phi \notin 3\text{SAT} & \implies f(\phi) \equiv (G, k) : G \text{ has no clique of size } \geq pk
\end{align*}
\]

However, thanks to the PCP Theorem 2.10, we have the freedom to use the NP-hardness of \( \text{gap}_{\alpha_0} - \text{MAX3SAT} \) to show NP-hardness of approximating \( \text{gap}_\rho - \text{MAXCLIQUE} \). That is, we now need a polynomial-time reduction \( g \) to do the following to 3CNF instances \( \phi \) (having \( m \) clauses):

\[
\phi \in 3\text{SAT} \implies g(\phi) \equiv (G, k) : G \text{ has a clique of size } \geq k
\]

Any assignment satisfies \( \leq \alpha_0 m \) clauses of \( \phi \) \( \implies g(\phi) \equiv (G, k) : G \text{ has no clique of size } \geq pk \)

Such reductions are called gap-preserving reductions, since it preserves the gap between satisfiable (YES) instances and far from satisfiable (NO) instances. The value of \( \rho \) that occurs in the inapproximability result, will, in general, depend upon the value of \( \alpha_0 \) and the nature of the reduction. In particular, a larger gap to start out with will invariably lead to a stronger hardness result.

Thus, the important question is obviously if the factor \( \alpha_0 \) in the above view is tight.

**Remark 2.12.** The FGLSS\[^{2.6}\] reduction can actually reduce a PCP verifier of the form \( \text{PCP}_{c,s}[r, q] \) to \( \text{gap}_{\alpha/c} - \text{MAXCLIQUE} \), with the reduction \( g \) running in time \( \text{poly}(2^{r+q}) \). That is, for any \( L \in \text{PCP}_{c,s}[r, q] \), if \( x \in L \), \( g(x) \in \text{YES} \), and if \( x \notin L \) then \( g(x) \in \text{NO} \) for \( \text{gap}_{\alpha/c} - \text{MAXCLIQUE} \). In particular, if the language \( L \) is NP-hard, with \( r, q = O(\log n) \), then we have MAX-CLIQUE is NP-hard to approximate to a factor of \( s/c \). Taking this into consideration, the above mentioned reduction from \( \text{gap}_{\alpha_0} - \text{MAX3SAT} \) is merely for illustration purposes.

However, rather than reducing instances of \( \text{gap}_{\alpha_0} - \text{MAX3SAT} \) to instances of problems we want to prove hardness results for, it is much more convenient to start from another place : the Label Cover problem.

2.2.2 Label Cover and 2-query PCPs

The Label Cover problem is a convenient starting point to produce reductions to problems that we want to show inapproximability results for. It abstracts out the interactive proof setting into a concrete gap-problem. We start by defining the problem :

**Definition 2.13.** (Label Cover Problem) A Label Cover instance \( I \) over the alphabet \( \Sigma = L \cup R \) (or label sets) consists of the pair \((G, \Pi)\), where :

- \( G = (U, V, E) \) is a bipartite graph with vertex sets \( U, V \) and edge set \( E \)
- \( \Pi = \{ \pi_e, e \in E | \pi_e : L \to R \} \) is a set of functions or constraints, one for each edge \( e = (u, v) \)
A labelling $A : U \to L$ and $B : V \to R$ is said to satisfy edge $(u, v)$ iff $\pi_c(A(u)) = A(v)$. The value of an instance $\text{val}(I)$ is the maximum fraction of satisfied edges.

For any $\epsilon \in (0, 1)$, $\text{gap}_\epsilon - \text{LC}_\Sigma$ is defined as the decision problem, given an instance $I$ over $\Sigma$, to decide which of the two classes $I$ belongs to:

$$\begin{align*}
\text{YES} & = \{ I \equiv (G, \Pi) : \text{val}(I) = 1 \} \\
\text{NO} & = \{ I \equiv (G, \Pi) : \text{val}(I) \leq \epsilon \}
\end{align*}$$

Remark. The form of the constraints used: that for any $l \in L$, there is a unique $r \in R$ such that $\pi(l) = r$ are said to have the projection property.

**Theorem 2.14.** $\exists \epsilon_0 \in (0, 1)$ and $\Sigma_0$ such that $\text{gap}_{\epsilon_0} - \text{LC}_{\Sigma_0}$ is NP-hard.

**Proof.** Following the obvious strategy, we show a gap-preserving reduction from 3SAT instances to Label Cover instances. Let

$$\phi = C_1 \land \ldots \land C_m$$

be a 3CNF formula on $n$ variables $x_1, \ldots, x_n$. The corresponding Label Cover instance $I(\phi)$ is the following:

- $U$ has one vertex for each clause $C_i$, so $|U| = m$.
- $V$ has one vertex for each variable $x_i$, so $|V| = n$.
- The edge set $E$ has an edge from clause $C$ to a variable $x$ iff $x$ or $\bar{x}$ appears in $C$.
- The right label set $R$ consists of satisfying assignments to literals appearing in the clauses: $R = \{0, 1\}^3 - \{(0, 0, 0)\}$. The left label set $L$ consists of assignments to individual labels, $L = \{0, 1\}$.
  - If $C = (x_1 \lor x_2 \lor \bar{x}_3)$ is assigned $(001)$ it corresponds to $x_1 \leftarrow 0, x_2 \leftarrow 0, x_3 \leftarrow 0$.
- The constraint on edge $(C_i, x_j)$ is that the assignment to variables in $C_i$ should be consistent with the assignment to $x_j$. That is, $A(C_i)|_{x_j} = B(x_j)$

*Proof that this is gap preserving: (Completeness)* Suppose $\phi \in 3\text{SAT}$. Then, we use the satisfying assignment to label $U$ and $V$. It is clear that all constraints are satisfied.

*(Soundness)* Suppose any assignment to $\phi$ satisfies $\leq \alpha$ fraction of the clauses. Pick any labelling $A, B$ of the label cover instance. The assignment $B$ can be consistent with at most an $\alpha$ fraction of the clause labellings in $A$. Since every clause has outdegree 3, we have:

$$\Pr_{e \in E}[e \text{ is satisfied}] = \sum_{C \in U} \Pr_{C}[C \text{ is picked}] \Pr_{x \in C_i}[\tau(x, C) \text{ is satisfied}]$$

$$= \sum_{C: \text{consistent with } B} \Pr_{C}[C \text{ is picked}] \Pr_{x \in C_i}[\tau(x, C) \text{ is satisfied}] + \sum_{C: \text{inconsistent with } B} \Pr_{C}[C \text{ is picked}] \Pr_{x \in C_i}[\tau(x, C) \text{ is satisfied}]$$

$$\leq \alpha \cdot \frac{1}{3} + (1 - \alpha) \cdot \frac{1}{3}$$

$$= \frac{1}{3} + \frac{2}{3} \epsilon$$

Invoking Theorem 2.10 for $\epsilon = \frac{1}{3} + \frac{2}{3} \alpha_0$, and $|\Sigma| = 9$, we get that $\text{gap}_\epsilon - \text{LC}_\Sigma$ is NP-hard.

**Relation between 2-query PCPs and Label Cover:** A Label Cover instance can be viewed as a 2-query PCP. More precisely, $\text{gap}_\epsilon - \text{LC}_\Sigma \in \text{PCP}^{\Sigma_0}_{\epsilon_0}[\log |E|, 2]$. Here, the proof to the verifier consists of the labellings $A, B$. On receiving an instance with a purported proof, the verifier chooses an edge constraint $(C, x)$ at random, and queries the proof for $A(C), B(x)$. The verification predicate simply checks if it is satisfied; the completeness and soundness properties of the PCP system is obvious.
2.2.3 The need for Parallel Repetition

While Theorem 2.14 can be used for reductions, it is not strong enough to prove strong hardness results. What we would like ideally is a result of the following form:

Claim 2.15. \( \forall \epsilon > 0, \text{gap}_c - LC_{\Sigma} \) is NP-hard for some \( \Sigma(\epsilon) \).

We note that the size of the alphabet set may be large, but is a constant depending on \( \epsilon \). From the previous paragraph, we can immediately conclude the following:

Corollary. (Of Claim 2.15) \( \forall \epsilon > 0, \exists \Sigma: \text{NP} \subseteq \text{PCP}_{1}^{\Sigma}(O(\log n), 2) \).

However, the task at hand is to produce a gap-preserving reduction from \( \text{gap}_c - LC_{\Sigma} \) to \( \text{gap}_c - LC_{\Sigma'} \). We can either view this in the 2-query PCP model, or in the Label Cover model. In either case, the main task is to get the soundness down to any given constant. For this, it is convenient to view Label Cover in the 2-Prover 1-Round game model.

From Label Cover to 2-Prover 1-Round Games

A Label Cover instance \( I \) in \( \text{gap}_c - LC_{\Sigma} \) can be viewed as a 2-Prover 1-Round game directly. The correspondence is as follows, in terms of the notation used in preceding discussions:

- Question set to \( P_1 \), \( (X')=U \), question set to \( P_2 \), \( (Y)=V \),
  - Joint distribution on question set is: uniform over pairs \((x, y) \in E\)
- Strategies for provers are labellings: \( f_1 = A, f_2 = B \)
- The verification predicate in the game is the constraint on edge \( e \): \( V(x, y, a, b) = 1 \iff \pi_{(x, y)}(a) = b \)
- The value of the label cover instance is the value of the corresponding game: \( \text{val}(I) = \text{val}(G) \)

In particular, NO instances of \( \text{gap}_c - LC_{\Sigma} \) result in games with value at most \( \epsilon \). Equivalently, using the same correspondence, any 2-Prover 1-Round game \( G \) with the projective property on verification constraints will correspond to an instance \( I(G) \) of Label Cover, with \( \text{val}(I(G)) = \text{val}(G) \). Henceforth, we can thus concentrate on the 2-Prover 1-Round game setting, and give a projection-preserving reduction that maps non-trivial game \( G \) to \( G' \), such that \( G' \) has low value. The mapping should preserve perfect completeness in the case of games with value 1.

Parallel Repetition

The parallel repetition of a 2-Prover 1-Round game described in Section 2.1.2 is a valid candidate for such a reduction. It is easy to see that perfect completeness and the projection property are preserved on moving from \( G \) to \( G^{\otimes n} \). However, as stated there, it is not obvious how the value of the game (equivalently, the soundness of the class of Label Cover instances produced) reduces.

Some trivial observations can be made, though, about the produced Label-Cover instances. The size of the instance increases exponentially in \( n \): the vertex set size blows up from \(|U| + |V|\) to \(|U|^n + |V|^n\); correspondingly the edge set goes from \(|E|\) to \(|E|^n\). The alphabet size moves from \(|\Sigma|\) to \(O(|\Sigma|^n)\).

2.3 Feige and Verbitsky’s Results

In [FV02], Feige and Verbitsky give arguments regarding what conditions a parallel repetition theorem, if proved, must satisfy. In particular, they explain why the dependence on the answer set size that arises in Theorem 2.3 is actually necessary. The main result in the paper is the following:

Theorem 2.16. There is a family of games \( \{G_k\}_{k \geq 2} \) such that \( \text{val}(G_k) \leq 3/4 \) and \( \text{val}(G_k^{\otimes n}) \geq 1/8 \) for \( n \leq k/4 \log k \). These games have \( |X| = |Y| = k \) and \( |A| = |B| = 2^{\Theta(k)} \).

Before giving the details of the proof of Theorem 2.16 we look at its consequences.
Corollary 2.17. For any constant $\alpha > 0$, there are games $G$ such that for large enough $n$, $\text{val}(G^{\otimes n}) > \text{val}(G)^\alpha$.

Proof. We use the family of games from the above theorem. Consider the game $G_k$; we will fix an appropriate $n, k$ later. For now, let $n = ck^4/\log k$; where $c \geq 2$. We divide the co-ordinates of the $n$-repeated game into $c$ blocks of $k/4 \log k$ length each, and play the game independently on each of these blocks. From the preceding result, we will have $\text{val}(G_k^{\otimes n}) \geq (\frac{1}{2})^c$. We now set $c$ such that

$$\left(\frac{1}{8}\right)^c > \left(\frac{3}{4}\right)^{\alpha n} \Rightarrow \frac{k}{\log k} > \frac{1}{\alpha \log 4k/3}$$

This can be done, by setting $k$ large enough; in particular, $k = \Omega(\log(\frac{1}{\alpha})).$ 

What the corollary implies is that we cannot have a theorem of the form $\text{val}(G^{\otimes n}) \leq \text{val}(G)^\alpha$ for $\alpha$ being a universal constant. That is, $\alpha$ must depend on the parameters of the game, say the question set size, or the answer set size. In fact, we can deduce more about $\alpha$, as stated in the following corollary:

Corollary 2.18. Let $\alpha$ be a function of $|X|$ and for every game $G$ and for every n, $\text{val}(G^{\otimes n}) \leq \text{val}(G)^\alpha$. Then $\alpha = O(|\log (|X|)/|X|))$. Similarly, if $\alpha$ is a function of $|A|$ with $\text{val}(G^{\otimes n}) \leq \text{val}(G)^\alpha$ for all $G, n$ then $\alpha = O(\log \log |A|/\log |A|))$.

Proof. We again look at the special family of games obtained from the theorem above. Using the proof of the preceding corollary, we saw that to ensure $\text{val}(G^{\otimes n}) \leq \text{val}(G)^\alpha$, we required that $\frac{k}{\log k} \leq \frac{1}{\alpha} \times \text{const}$. But for the family of games considered, $k = |X| = \log |A|$. Substituting these values gives us the result.

We now turn to sketch the proof of the main result, Theorem 2.16.

Proof. (of Thm 2.16 Sketch) Fix any $n$; we will define games $G_k$ for $n \leq \frac{k}{\log k}$. The game $G_k$ will have the following parameters:

- $X = \{x_1, \ldots, x_k\}$, $\mathcal{Y} = \{y_1, \ldots, y_k\}$, the distribution $D$ is uniform on $X \times \mathcal{Y}$
- Answers for the players are $A = [n] \times X^n$, $B = [n] \times \mathcal{Y}^n$
- The predicate depends upon a graph $G_k^n$, which we shall define shortly. For brevity, denote $\bar{x} \equiv x_1 \ldots x_n$ and $\bar{y} \equiv y_1 \ldots y_n$;

$$V(x, y, (l \circ x_1 \ldots x_n), (m \circ y_1 \ldots y_n)) = 1 \iff l = m \land x_1 = x \land y_m = y \land (\bar{x}, \bar{y}) \in E(G_k^n)$$

$G_k^n$ is a bipartite graph with vertex sets $X^n, \mathcal{Y}^n$. There are $k^2n/8$ edges, arranged such that any $k \times k$ induced subgraph has atmost $k^2/2$ edges. First off, it is necessary to show that such a graph exists:

Claim. $G_k^n$ as above exists, for $n \leq \frac{k}{4\log k}$.

Proof. The proof is by a probabilistic argument: generate the graph by picking every possible edge at random with probability $\frac{1}{2}$. Then, $\mathbb{E}[|E|] = \frac{k^2n}{8}$, and $\Pr[|E| \geq \frac{k^2n}{8}] \approx \frac{1}{2}$. Now, picking any $k \times k$ subgraph $H$, $\Pr[|E(H)| \geq \frac{k^2}{2}] \leq \exp(-\frac{9}{32k^2})$. By the union bound, $\Pr[|E(H)| \geq \frac{k^2}{2}] \leq k^{2n} \exp(-\text{const} \times k^2)$. Choosing $n \leq k/4 \log k$, this quantity is less than $\frac{1}{2}$ for large enough $k$. Combining with the previous estimate of the bad event that the number of edges in $G$ drops below $k^2n/8$, we infer that the claimed graph exists with positive probability.

Claim. $\text{val}(G_k^{\otimes n}) \geq \frac{1}{8}$
Proof. Suppose P1 receives $\bar{x} = x_1 \ldots x_n$, and P2 receives $\bar{y} = y_1 \ldots y_n$. P1 answers back on co-ordinate $i$ with $a_i = (i \circ x_1 \ldots x_n)$ and P2 answers with $b_i = (i \circ y_1 \ldots y_n)$. They pass iff $(\bar{x}, \bar{y}) \in E(G^n_k)$, and since all questions are equally probable, this happens with probability at least $\frac{k^{2n/3}}{2^n} = \frac{1}{2}$. □

Claim. $\text{val}(G_k) \leq \frac{3}{4}$

Proof. Consider strategies for the players where P1 answers $x_i \in X$ with $l_i \circ \bar{x}^i$ and P2 answers $y_i \in Y$ with $m_i \circ \bar{y}^i$. Define the acceptance graph $G_{\text{Acc}}$ for this strategy to be the bipartite graph on vertex set $X, Y$ with edges $(x_i, y_j)$ iff the players succeed on receiving that question pair. We shall prove that $G_{\text{Acc}}$ cannot have too many edges. First, we make some simple observations:

Proposition. $l_i = m_j$ if $x_i$ and $y_j$ are in the same connected component in $G_{\text{Acc}}$.

Proof. Obvious. □

Proposition. Any connected component of $G_{\text{Acc}}$ has atmost $\frac{k^2}{2}$ edges

Proof. “Lift” the edges in the connected component of $G_{\text{Acc}}$ to edges in $G^n_k$ by using the answers of the players on the corresponding questions. For instance, $(x_i, y_j)$ would go to $(\bar{x}^i, \bar{y}^j)$ in $G^n_k$. The edge $\bar{x}^i, \bar{y}^j$ exists in $G^n_k$, since the verifier accepts on $(x_i, y_j)$. Further, it is easy to see that no two edges of $G_{\text{Acc}}$ map to the same edge in $G^n_k$. So, we have a connected component corresponding to part of a $k \times k$ induced subgraph in $G^n_k$, which implies that it can have no more than $\frac{k^2}{2}$ edges. □

Finally, we show that $G_{\text{Acc}}$ can have atmost $3k^2/4$ edges. Divide $G_{\text{Acc}}$ into connected components $\{C_i\}_{i \in [t]}$, where each $C_i$ occupies $a_i$ vertices on the left and $b_i$ vertices on the right. Arrange $C_i$ in order of decreasing $b_i$, and consider $C_1, C_2$. The number of edges not in $G_{\text{Acc}}$ between $C_1, C_2$ is $a_1b_2 + a_2b_1 \leq a_2b_2$. But $a_2b_2$ is the maximum number of edges in $C_2$. Hence, for every edge that $C_2$ has, there is a corresponding edge in $G_{\text{Acc}}$ that is absent! This argument works for every component $C_i$ for $i \geq 2$. Thus, the total number of edges in $G_{\text{Acc}}$ is atmost:

$$|E| \leq |E(C_1)| + \frac{k^2 - |E(C_1)|}{2}$$

$$= \frac{k^2}{2} + \frac{|E(C_1)|}{2}$$

$$\leq \frac{3k^2}{4}$$

This shows that the acceptance probability is atmost $\frac{3}{4}$. □

2.4 Feige and Kilian’s Repetition theorem

Now that we have a good idea of how a parallel repetition theorem should work, we mention a result of Feige and Kilian[FK00]. The theorem concerns a class of games they call miss-match games, defined as follows:

Definition 2.19. (Miss-Match games) A miss-match game $G_{\text{MM}}$ is a 2-Prover 1-Round game, with the following form:

- the question set for P1 is of the form $X = X_1 \times X_2 \times \{\lambda\}$
- the question set for P2 is $Y = X_1 \cup X_2 \cup \{\lambda\}$
- the distribution over the questions is: arbitrary (some D) over $X_1 \times X_2$ for P1, and P2’s question is one of the co-ordinates of P1, i.e. one of $x_1, x_2, \lambda$ with uniform probability, when P1 receives $(x_1, x_2, \lambda)$
• answer sets are: $A = A_1 \times A_2 \times \{\lambda\}$ and $B = A_1 \cup A_2 \cup \{\lambda\}$
• the acceptance predicate $V = V_1 \lor V_2$, where $V_1 : X_1 \times X_2 \times A_1 \times A_2 \rightarrow \{0,1\}$ and $V_2 : X_1 \times X_2 \times A \times B \rightarrow \{0,1\}$ is a projection constraint that checks for consistency: $V_2((x_1 x_2 \lambda), y, (a_1 a_2 \lambda), b) = 1$ iff $y = x_1 \land b = a_1$ or $y = x_2 \land b = a_2$ or $y = \lambda \land b = \lambda$.

Here, $\lambda$ may be interpreted as a special character, which if received by any of the players, is called a “miss”. If the question to $P_2$ is $\lambda$, the round is called a miss. If $P_2$ receives a question from $X_1 \cup X_2$, we call it a match round.

Miss-match games are not very far from general games, as the next theorem proves:

**Theorem 2.20.** For every game $G$, we can define a miss-match game $G_{MM}$ that satisfies $\text{val}(G_{MM}) \leq \frac{\text{val}(G) + 2}{3}$

**Proof.** Set the parameters of $G_{MM}(A', Y', D', A', B', V')$ in the natural way from $G(X, Y, D, A, B, V)$

• $A' = X \times Y \times \{\lambda\}$; $Y' = X \cup Y \cup \{\lambda\}$, $D'|X \times Y' \equiv D$, and $P_2$’s question is then chosen by the miss-match structure.
• $A' = A \times B \times \{\lambda\}$, $B' = A \cup B \cup \{\lambda\}$
• $V'(x', y', a', b') = V(x, y, a, b) \land V_2$, where $V_2$ is as in Definition 2.19.

We note that the above transformation preserves the projection property, if the original game has it. Let the players pick any strategy $f_1, f_2$ for $G_{MM}$. For some fraction of the questions in $\mathcal{X}'$, $P_1$ answers inconsistently with $P_2$ on at least one of the questions. Call this set $X_1$, and let it have weight $\alpha$ under $D'$. Let $W$ denote the event that the players win $G_{MM}$, and $E$ be the event that $P_1$ gets a question from $X_1$. Then:

$$P_D[W] = \Pr[E] \Pr[V_1 \land V_2 | E] + \Pr[\overline{E}] \Pr[V_1 \land V_2 | \overline{E}]$$

$$\leq \alpha \Pr[V_2 | E] + \Pr[V_1 \land E]$$

$$\leq \frac{\alpha^2}{3} + \min\{\Pr[V_1], \Pr[E]\}$$

$$= \frac{\alpha^2}{3} + \min\{\text{val}(G), 1 - \alpha\}$$

$$\leq \frac{2\alpha}{3} + \frac{\text{val}(G) + 2(1 - \alpha)}{3}$$

$$= \frac{2 + \text{val}(G)}{3}$$

The main result regarding miss-match games is the following:

**Theorem 2.21.** (Parallel Repetition of Miss-Match Games, [FK92]) Let $G_{MM}$ be any miss-match game derived from a game $G$ with $\text{val}(G) = 1 - \epsilon$. Then $\text{val}(G_{MM}^n) \leq \delta$ whenever $n \geq (\frac{c}{\epsilon})^c$, where $c \geq 0$ is a universal constant independent of $G, \epsilon, n, \delta$.

Thus, the theorem requires $\text{poly}(\frac{1}{\epsilon})$ repetitions to get down the value to below $\delta$. An interesting point to note is that this result manages to bypass any dependence on the original game parameters, other than $\epsilon$, in contrast to what Section 2.3 predicted. The miss-match structure somehow avoids this requirement.

We omit the proof of Theorem 2.21 for lack of space. However we mention the intuition regarding why miss rounds facilitate error reduction under parallel repetition. In a game with $m$ miss rounds out of $n$, the prover $P_2$ can correlate his answers only on $n - m$ of the $n$ co-ordinates. On the miss rounds, he has no idea of what $P_1$ received. On the other side, however, $P_1$ ’s answers on the questions are correlated across all co-ordinates. Thus, $P_1$ and $P_2$ cannot really co-ordinate their strategies due to the presence of the miss rounds, at least not as well as they would have done in the original game.
2.4.1 Comparison of Parameters with Raz’s theorem

Raz’s parallel repetition theorem had the error rate going down exponentially: if $\text{val}(G) \leq 1 - \epsilon \implies \text{val}(G^\otimes n) \leq (1 - c\epsilon)^d n^{-c\log n}$, where $c, d$ are absolute constants. Thus, to get the error rate down to below $\delta$, one needed $O(\log \frac{1}{\delta})$ repetitions, ignoring the dependence on other parameters of the game.

Recall the the Label Cover view: Suppose we wanted to show that $\text{gap} gap_{1,\delta} - LC_{\Sigma}$ is hard for some $\Sigma$, we used a transformation on label cover instances obtained by a gap-preserving polynomial-time reduction from MAX-3SAT. The transformation, as we see, can use either Feige-Kilian’s or Raz’s repetition theorem. However, the size of the produced Label Cover instances, and hence the running time of the reduction, depends on the number of rounds in parallel repetition required to get the soundness down to $\delta$. The following table compares these parameters (assuming $\epsilon_0, \Sigma_0$ in gap $\text{gap}_{1,\epsilon_0} - LC_{\Sigma_0}$ are constant):

| Theorem Used | Number of repetitions | Number of vertices | Alphabet size ($|\Sigma|$) |
|--------------|-----------------------|-------------------|--------------------------|
| Feige-Kilian 2.21 | poly($\frac{1}{\delta}$) | $|V|^\text{poly}(\frac{1}{\delta})$ | $2^{\text{poly}(\frac{1}{\delta})}$ |
| Raz 2.5        | $O(\log(\frac{1}{\delta}))$ | $|V|^{O(\log \frac{1}{\delta})}$ | poly($\frac{1}{\delta}$) |

In the reduction from MAX-3SAT instances to Label Cover instances, we see that the degree of the left vertices is a constant, so the running time of the reduction is $O(\text{vertex set size})$.

In fact, there is no need to assume $\delta$ to be an absolute constant. We could ask for, say, $\delta = \frac{1}{\log^3 n}$, where $n$ is the size of the $3CNF$ formula we reduce from. In this case, the running time of Raz’s reduction is $n^{O(\log \log n)}$, while Feige-Kilian requires $n^{O(\text{polylog} n)}$. Thus, using Raz’s theorem, we conclude the following stronger result:

**Theorem.** $\forall n, \exists \Sigma \text{ gap}_{1,\frac{1}{\log^3 n}} - LC_{\Sigma}$ is not in $P$ unless $NP \subseteq DTIME(n^{O(\log \log n)})$

With this, we conclude the discussion regarding the role of parallel repetition of 2-Prover 1-Round games in hardness of approximation results. We move on to (a simplified version) of the proof of Raz’s Parallel Repetition result.
3 Proof of the Parallel Repetition Theorem

In this section, we give a proof of Raz’s Parallel Repetition theorem. The version of the proof here is taken from [Hol07, Rao08], the notation closely follows that used in the latter. Rao [Rao08] also gives a stronger version of the theorem applicable to projection games; we also discuss this. Finally we move on to a concentration result mentioned in the same paper.

Notation:
The notation for 2-Prover 1-Round games used is consistent with the one introduced in Section 2.1. For convenience, we will denote the size of the answer set of prover $P_1$ by $2^c = |A|$.

For a random variables $X, Y, Z$ over the space $\Omega$, $\{X\}$ will stand for the distribution of $X$, $\{XY\}$ will be the joint distribution of $X$ and $Y$ over $\Omega$, and $\{XY|Z\}$ will denote the conditional distribution of $X, Y$ given $Z$; i.e. $\{X = x, Y = y|Z = z\} = \Pr[X = x \land Y = y|Z = z]$. Given an event $E \subseteq \Omega$, $\{X|E\} \triangleq \Pr[X|E]$. Hence, $D \triangleq \{Z|E\}\{XY|Z\}$ will be the distribution over triples $(x, y, z) \in \mathbb{R}^3$ with $\Pr_D(X, Y, Z = (x, y, z)) = \Pr[Z = z|E]\Pr[X = x \land Y = y|Z = z]$. Finally, we let $\|X - Y\|_1$ be the $L_1$-distance between variables $X, Y$ (equivalently, their distributions).

3.1 The Theorem and Proof Methodology

The main theorem we prove is the following:

Theorem 3.1. (Parallel repetition for general games [Hol07, Rao08]) If $G$ is a game with $\text{val}(G) = 1 - \epsilon$, then the value of the $n$-repeated game satisfies

$$\text{val}(G^\otimes n) \leq (1 - \epsilon)^{\frac{\epsilon^2n}{\alpha^2 n}}$$

where $\alpha > 0$ is a universal constant.

Proof Method: Let $W$ denote the event that the provers win on all co-ordinates, and $W_S$ denote the event that they win on a subset $S \subseteq [n]$ of the co-ordinates. The basic idea behind the proof is to move inductively on the co-ordinates, showing that at every point, there is still a hard-to-win co-ordinate left over. More precisely, since

$$\Pr[W_{S \cup \{i\}}] = \Pr[W_i|W_S] \Pr[W_S]$$

- if $\Pr[W_S]$ is low enough, we are done already
- if $\Pr[W_S]$ is high, then most other co-ordinates are biased towards losing: $E_{j \notin S}[\Pr[W_j|W_S]]$ is low

How we go about arguing the second point will occupy most of the proof. The method used for this will be the following: a random co-ordinate $j \notin S$ of the $n$-round game, conditioned on $W_S$, does not look too different from a single round game. So in effect, the provers can embed $G$ into a co-ordinate $j$ of $G^\otimes n$, by generating the rest of the questions according to a distribution close to $X^n|X_j = X \land Y_j = Y \land W_S$, with and hence should be able to win $G$ with probability close to $\Pr[W_j|W_S]$. The issue, however, is that each player does not know the other’s question and they have to do this without communication; even though the generated questions are highly correlated and depend on questions received by both players in $G$ due to the conditioning on $W_S$.

To do this effectively, they use the method of correlated sampling introduced in [Hol07]. Now, onto the proof.

3.2 The Proof

The main lemma we will prove is the following:
Lemma 3.2. (Main Lemma) For all $S \subseteq [n]$ , with $|S| = k$ we have that for the game $G \otimes^n$

$$E_{i\in S} [\Pr[W_i|W_S]] \leq \text{val}(G) + 15 \sqrt{\frac{k \log |A| + \log \frac{1}{\Pr[W_S]}}{(n-k)}}$$

$$= \text{val}(G) + 15 \sqrt{\frac{kc + \log \frac{1}{\Pr[W_S]}}{(n-k)}}$$

In particular, if $\Pr[W_S] \geq 2^{-\gamma(n-k)+kc}$ for some $\gamma \geq 0$, then $E_{i\in S} [\Pr[W_i|W_S]] \leq (1 - \epsilon + 15\gamma)$.

Let us see how this gives us the main theorem:

Proof. (Of Theorem 3.1 using Lemma 3.2) We proceed by induction on the size of the subset $S$ to prove the following:

Claim. There is a set $S \subseteq [n]$ with $|S| = k$ such that $\Pr[W_S] \leq (1 - \epsilon/2)^k$, where $k$ is “not too large”. (We will quantify this in what follows)

Proof. (Base case) For $|S| = 0$, $S = \emptyset$, and for all $i \in [n]$, $\Pr[W_i|W_S] = \Pr[W_i] \leq \text{val}(G) = 1 - \epsilon \leq 1 - \epsilon/2$.

This is because in $G$, the two provers can simply use shared randomness to produce the questions $X_j, Y_j$ in each co-ordinate independently using the distribution $\{XY\}$ and answer according to the strategy for $G \otimes^n$. This works since the conditioning is on the whole space, and does not introduce dependencies across co-ordinates.

(Inductive Hypothesis): There is a $S$ with $|S| = k$ and $\Pr[W_S] \leq (1 - \epsilon/2)^k$.

(Induction): To prove that there is some $i \notin S$ such that $\Pr[W_{S\cup\{i\}}] \leq (1 - \epsilon/2)^{k+1}$. Suppose that $\Pr[W_S] \leq (1 - \epsilon/2)^{k+1}$, then we are done already. Suppose not, and that $\Pr[W_S] \geq (1 - \epsilon/2)^{k+1} \geq (1/2)^{k+1}$. Then invoking Lemma 3.2 gives us the following:

$$\exists i \notin S : \Pr[W_i|W_S] \leq (1 - \epsilon) + 15 \sqrt{\frac{kc + \log \frac{1}{\Pr[W_S]}}{n-k}}$$

$$\leq (1 - \epsilon) + 15 \sqrt{\frac{kc + k + 1}{n-k}}$$

$$\leq (1 - \epsilon) + 15 \sqrt{\frac{2kc}{n-k}}$$

For the induction to go through, we would like to keep the value of this to below $1 - \epsilon/2$. This would be satisfied if :

$$15 \sqrt{\frac{2kc}{n-k}} \leq \frac{\epsilon}{2}$$

$$\iff k \leq \frac{\epsilon^2 (n-k)}{9000c}$$

$$\iff k \leq \frac{\epsilon^2 n}{18000c}$$

Thus, for $k$ up to the above value, we have that $\exists i : \Pr[W_i|W_S] \leq (1 - \epsilon/2)$. We can include this $i$ into the set $S$ and continue. This finishes the claim.

Using the above theorem, the proof of the main theorem follows immediately :

$$\Pr[W_n] \leq \Pr[W_S]$$

$$\leq (1 - \frac{\epsilon}{2})^{\epsilon^2 n / 18000c}$$

$$\leq (1 - \epsilon)^{\alpha \epsilon^2 n/c}$$ for some constant $\alpha$
What Lemma 3.2 basically says is that the probability of winning a random co-ordinate outside $S$ is not too large compared to the one co-ordinate game. We show this by using a strategy for $G^{\otimes n}$ in $G$, with the players in $G$ almost simulating the $n$-repeated game conditioned on $W_S$. We now describe how this is done.

### 3.2.1 Local Embeddings

Let us make the notion of embedding precise. In order to motivate this definition, consider the situation $P1$ and $P2$ are in: $P1$ receives $X_0$, and needs to come up with $X = (X_1 \ldots X_n)$, such that $X_i = X_0$ using simply a shared random source $R \leftarrow \mathcal{R}$. Similarly, $P2$ gets $Y_0$ and needs to come up with $Y = (Y_1 \ldots Y_n)$ such that $Y_j = Y_0$ using the same $R \leftarrow \mathcal{R}$. The final distribution they get should be close to $X^nY^n|W_S$. In effect, they need to use the shared source to correlate their local computations.

**Definition 3.3.** (Local embeddings) Let $X, Y, T$ be random variables, and $X_0, Y_0$ be over the same same space as $X, Y$ respectively. Then, we say that $X, Y$ can be $\epsilon$-locally embedded in $XS, YT$ with $X, Y = X_0, Y_0$ if there is a random variable $R \in \mathcal{R}$ and functions $F_1 : \mathcal{X} \times \mathcal{R} \rightarrow \mathcal{S}$ and $F_2 : \mathcal{Y} \times \mathcal{R} \rightarrow \mathcal{T}$ such that:

$$\|\{XSYT\} - \{X_0Y_0\}\{F_1F_2XY\}\|_1 \leq \epsilon$$

We can thus state the problem as follows:

- Find an appropriate probability space $\mathcal{R}$ that both $P1$ and $P2$ sample from.
- Effectively, this sampling for $R$ should reduce the problem to a local one: The players can now compute local functions to produce the tuple of questions in their half.

The second point can be interpreted as the fact that conditioned on any fixed value of $R, X_i, Y_i$ and $W_S$, $(X^nY^n)$ becomes a product distribution:

$$\{X^nY^n|R, X_i, Y_i \land W_S\} = \{X^n|R, X_i, Y_i \land W_S\}\{Y^n|R, X_i, Y_i \land W_S\}$$

We will draw $R$ from a part of the space $\mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{A}^n \times \mathcal{B}^n$ (distribution over the provers winning strategy for $G^{\otimes n}$), such that it breaks dependence between $X^n$ and $Y^n$. Since variables from this space, conditioned on $W_S$, would depend on both $X_i$ and $Y_i$, the distribution the players must ideally draw $R$ from is $\{R|X_i = x \land Y_i = y \land W_S\}$. However, $P1$ does not know $y$. The following method describes a way to bypass this requirement.

### 3.2.2 Correlated Sampling

The following fact will be useful in what follows:

**Fact 3.4.** If $X, X'$ are random variables over $\mathcal{X}$ in the same probability space with $\Pr[X \neq X'] \leq \epsilon$, then $\|X - X'\|_1 \leq \epsilon$.

**Theorem 3.5.** (Correlated Sampling, Holenstein[Hol07]) Let $X, Y$ be random variables. If there exist $\epsilon_1, \epsilon_2 \in [0, 1]$ such that

$$\|\{XSY\} - \{XY\}\{S|X\}\|_1 \leq \epsilon_1$$

$$\|\{XSY\} - \{XY\}\{S|Y\}\|_1 \leq \epsilon_2$$

Then, $(X, Y)$ is $1 - 2\epsilon_1 - 2\epsilon_2$-locally embeddable in $(XS,YS)$.

Here, think of $S$ as the shared random string they draw. What this says is that if we could somehow ensure that the information required to produce $S$ is contained almost fully in $X_i$ alone, and $Y_i$ alone, then it possible for the provers to sample $S$ without communication. We now give the proof of this key result.
Proof. Think of Alice as having $X_0$ and producing $X_0S$ and Bob as having $Y_0$ and producing $Y_0S$.
Consider the following method of producing sample $S$, given that $X = x$ and $Y = y$:

1. At iteration $i$, pick a random point $(p_i, s_i)$ in $[0, 1] \times S$.
2. If Alice does not yet have an output, and $p_i \leq \Pr\{S = s_i | X = x\}$, Alice outputs $S_A = s_i$ (this is Alice’s local function $F_A$).
3. If Bob does not yet have an output and $p_i \leq \Pr\{S = s_i | Y = y\}$, Bob outputs $S_B = s_i$ (this is Bob’s local function $F_B$).
4. Repeat step 1 until both players have an output. (With probability 1, this procedure terminates)

Remark. Here, the shared random space is the uniform distribution on $(S \times [0, 1])^\infty$, since at every step, we choose a point in $S \times [0, 1]$.

For the sake of analysis, consider a third player Dexter, also participating in the above process, who picks the first $i$ such that $p_i \leq \Pr\{S = s_i | X = x \land Y = y\}$ and outputs $S_D = s_i$. It is clear that $S_A \sim \{S|X\}$, $S_B \sim \{S|Y\}$ and $S_D \sim \{S|XY\}$. Dexter’s distribution of variables $\{XYS\}$ is our desired distribution. Consider the event $S_A \neq S_D$. The probability of this event happening is:

$$\Pr\{S_A \neq S_D\} = 1 - \frac{\sum_x \min(\Pr\{S = s | X = x\}, \Pr\{S = s | X = x \land Y = y\})}{\sum_x \max(\Pr\{S = s | X = x\}, \Pr\{S = s | X = x \land Y = y\})}$$

Similarly, $\Pr\{S_B \neq S_D\} \leq 2\|\{S|Y = y\} - \{S|X = x \land Y = y\}\|_1 \implies \Pr\{S_A = S_B = S_D\} \geq 1 - 2\|\{S|X = x\} - \{S|X = x \land Y = y\}\|_1 - 2\|\{S|Y = y\} - \{S|X = x \land Y = y\}\|_1$. Averaging over $\{XY\}$, we get that $\Pr\{S_A = S_B = S_D\} \geq 1 - 2\epsilon_1 - 2\epsilon_2$. Since $\{XYS\} \equiv \{XYSD\}$, we use Fact 3.4 to finish with:

$$\|\{XYS\} - \{XY\}\{SA|XY\}\|_1 \leq 2\epsilon_1 + 2\epsilon_2$$

The following Corollary is more useful in our case:

Corollary 3.6. If $(X_0, Y_0)$ is over the same space as $(X, Y)$ and

$$\|\{XY\} - \{X_0Y_0\}|S|X\|_1 \leq \epsilon_1$$

$$\|\{XY\} - \{X_0Y_0\}|S|Y\|_1 \leq \epsilon_2$$

Then $(X_0Y_0) \in 1 - 2\epsilon_1 - 2\epsilon_2 - \min(\epsilon_1, \epsilon_2)$-locally embeddable in $(XS|YS)$ with $(X, Y) = (X_0, Y_0)$

Proof. The proof follows by noting that the given conditions, on marginalizing over the space $S$ imply that $\|\{XY\} - \{X_0Y_0\}\|_1 \leq \min(\epsilon_1, \epsilon_2) \equiv \epsilon_{12}$. Then there is a joint probability space over $XYX_0Y_0$ such that the marginal distributions remain the same, but $\Pr\{XY \neq X_0Y_0\} \leq \epsilon_{12}$. The players use the same functions $s_A, s_B$ as in the previous theorem. We again invoke Fact 3.4 on the variables $\{XYS\}$ against $\{X_0Y_0SA|SB\}$ to conclude the proof.

3.2.3 Proof of Main Lemma

Let, without loss of generality, $S = \{n-k+1, \ldots, n\}$ ($|S| = k$). We define some of the random variables used in this part of the proof:

- $A = A_{n-k+1} \ldots A_n$ is the answers $P1$ gives in $S$. $B = B_{n-k+1} \ldots B_n$ are the answers $P2$ gives in $S$.
- $Q = X_{n-k+1} \ldots X_nY_{n-k+1} \ldots Y_n$ are the questions in $S$
• \( V = V_1 \ldots V_{n-k} \) are uniformly random bits

• For \( i \notin S \), \( T_i \stackrel{def}{=} \begin{cases} X_i & \text{if } V_i = 1 \\ Y_i & \text{if } V_i = 0 \end{cases} \), i.e. it is a random question in every co-ordinate

• For \( i \notin S \), \( U_i \) denotes the opposite questions to \( T_i \), i.e. \( U_i = X_i \) if \( V_i = 0 \) and \( U_i = Y_i \) if \( V_i = 1 \)

With a view towards breaking the dependence between \( X^n, Y^n \) (conditioned on \( W_S \)), we introduce the following crucial definition of the random variable \( \Lambda \):

• To break the dependence between \( X_i \) and \( Y_i \) for \( i \notin S \), \( \Lambda \) will contain \( T_1 \ldots T_{n-k} V_1 \ldots V_{n-k} \) (i.e. a random question from each co-ordinate)

• To break the dependence between \( X_i \) and \( Y_j \) for \( i \neq j \) (this was introduced due to the conditioning on \( W_S \)), \( \Lambda \) contains \( Q, A \)

For convenience, club together the questions in \( \Lambda \) into \( R = QTV \), so that \( \Lambda = AR \). When considering embedding into co-ordinate \( j \notin S \), we will be concerned with \( \Lambda^{-j} \equiv AR^{-j} \), where \( R^{-j} \equiv QT^{-j}V^{-j} \) is \( R \) without variables \( T_j, V_j \). We note the following independences induced by the definition of \( \Lambda^{-j} \):

\[
\begin{align*}
\{X^n Y^n & \Lambda^{-j}|W_S\} = \{X_i Y_j|W_S\}\{\Lambda^{-j}|X_i Y_j \wedge W_S\}\{X^n Y^n|\Lambda^{-j} X_i Y_j \wedge W_S\} \\
& = \{X_i Y_j|W_S\}\{\Lambda^{-j}|X_i Y_j \wedge W_S\}\{X^n|\Lambda^{-j} X_i Y_j \wedge W_S\}\{Y^n|\Lambda^{-j} X_i Y_j \wedge W_S\} \\
& = \{X_i Y_j|W_S\}\{\Lambda^{-j}|X_i Y_j \wedge W_S\}\{X^n|\Lambda^{-j} X_i Y_j \wedge W_S\}\{Y^n|\Lambda^{-j} Y_j \wedge W_S\}
\end{align*}
\]

(3.5)

close to \( \{XY\} \) generated by \( P_1 \)

correlated sampling \( \Lambda \)

generated by \( P_2 \)

The first step is to show that conditioned on \( W_S \), a dense event, the first term is close to the original single co-ordinate distribution. To do this, we use the following Lemma (which we do not prove):

**Lemma 3.7.** Let \( M_1 \ldots M_l \) be independent random variables. Then, for any event \( E \), we have :

\[
E_{\ell \in [l]}||\{M_{\ell}|E\} - \{M_{\ell}\}||_1 \leq \sqrt{\frac{\log(\frac{1}{\epsilon})}{l}}
\]

Thus, if \( W \) is dense enough, on an average, conditioning does not change the distributions of independent random variables by too much. The above lemma applied on the single variable \( \{X_i, Y_i\} \) yields :

**Corollary 3.8.** \( E_{\ell \in S}||\{X_i, Y_i|W_S\} - \{X_i Y_i|W_S\}||_1 \leq \sqrt{\frac{\log(\frac{1}{\epsilon})}{|W_S|}} \)

The heart of the proof involves showing the following :

\[
\{\Lambda^{-j}|X_j = x \wedge W_S\} \approx \{\Lambda^{-j}|X_j = x \wedge Y_j = y \wedge W_S\} \approx \{\Lambda^{-j}|Y_j = y \wedge W_S\}
\]

To do this, we will need the following Lemma that extends Lemma 3.7 to variables that are independent conditioned on another variable (as it is in our case)

**Lemma 3.9.** ([Hol02]) Let \( P \) in \( \mathcal{P} \), \( M_1 \ldots M_l \) in \( \mathcal{M} \), \( N \) in \( \mathcal{N} \) be random variables such that \( \{PM^lN\} = \{P\} \{M_1|E\} \ldots \{M_l|P\}\{V|PM^l\} \) and \( E \) be an event. Then,

\[
E_{\ell \in |l|}||\{NP|E\} - \{N\}|M_\ell|P\}||_1 \leq \sqrt{\frac{\log |N^*| + \log(\frac{1}{\epsilon})}{|E|}}
\]

where \( N^* = \text{supp}(N|E) \)

20
We invoke this Lemma, with $P \leftarrow R$, $M_i \leftarrow U_i$, $N \leftarrow A$ and $E \leftarrow W_S$ (since questions across co-ordinates are independent even when conditioned on $R = r$) to get:

$$
\mathbb{E}_{i \in S}[\|\{AR|W_S\} - \{AR|W_S\}U|R\|_1] \leq \sqrt{\frac{\log |A|^k + \log \left(\frac{\text{supp}(AR|W)}{\text{supp}(AR|W_S)}\right)}{n-k}} = \delta
$$

(3.6)

Where we set $\delta := \sqrt{\frac{k+\log(\frac{|A|}{|A^k|})}{n-k}}$ for convenience, and used the fact that $\text{supp}(A^k) \leq \text{supp}(A^k)$. We now use the following fact about conditioning random variables:

**Fact 3.10.** Let $Z_1, Z_2$ be random variables in $\mathcal{Z}$. Let $S \subseteq \mathcal{Z}$ be such that $\Pr[Z_1 \in S] = \Pr[Z_2 \in S] = \frac{1}{2}$. Then we have

$$
\|\{Z_1^i|Z_1 \in S\} - \{Z_2^i|Z_2 \in S\}\|_1 \leq 2\|\{Z_1\} - \{Z_2\}\|_1
$$

Conditioning each term in the expectation on the probability $\frac{1}{2}$ set $V_i = 0$ (we note that $V_i$ is independent of $W_S$), we get that:

$$
\mathbb{E}_{i \in S}[\|\{AR_j^i|W_S \land V_1 = 0\} - \{AR_j^i|W_S \land V_1 = 0\}U|R \land V_1 = 0\|_1] \leq 2\delta
$$

$$
\implies \mathbb{E}_{i \in S}[\|\{AR_j^{-i}X_1^i Y_1^i|W_S\} - \{AR_j^{-i}X_1^i Y_1^i|W_S\}\|_1] \leq 2\delta
$$

(removing unnecessary terms)

$$
\implies \mathbb{E}_{i \in S}[\|\{\Lambda^{-i}X_1^i Y_1^i|W_S\} - \{X_1^i Y_1^i\}\|_1] \leq 2\delta
$$

(rearranging terms)

But, we know that $\mathbb{E}_{i \in S}[\|\{Y_i^i\} - \|\{Y_i|W_S\}\|_1] \leq \sqrt{\log(\frac{\text{supp}(AR|W_S)}{\text{supp}(AR|W_S)})} \leq \delta$, implying can use $\{Y_i\}$ as a surrogate for $\{Y_i|W_S\}$ above, while losing only $\delta$, giving

$$
\mathbb{E}_{i \in S}[\|\{\Lambda^{-i}X_1^i Y_1^i|W_S\} - \{X_1^i Y_1^i\}\|_1]\leq 3\delta
$$

Symmetric reasoning yields that

$$
\mathbb{E}_{i \in S}[\|\{\Lambda^{-i}X_1^i Y_1^i|W_S\} - \{X_1^i Y_1^i\}\|_1]\leq 3\delta
$$

The above two conditions imply that on average across co-ordinates, the provers can use Theorem 3.5 to generate variables that are $2(3\delta) + 2(3\delta) + 3\delta = 15\delta$ close to the distribution $\{\Lambda^{-i}X_1^i Y_1^i|W_S\}$ without communicating. Using the independence properties that $\Lambda^{-i}$ induces (Eqn. (3.5)), the players, on receiving $X = x, Y = y$ in $\mathcal{G}$, can generate the remaining questions without communicating, and produce a joint distribution that is $15\delta$-close to $\{X^n Y^n\}^{-i}|W_S\}$ while having $X_i = x, Y_i = y$. They then respond back with the strategy for $\mathcal{G}^{\otimes n}$, to win the single round game with probability at least $\Pr[W_i|W_S] - 15\delta$ (on average across $i$). More precisely,

$$
\mathbb{E}_{i \in S}[\Pr[W_i|W_S]] \leq \text{val}(\mathcal{G}) + 15\delta
$$

We have thus proved Lemma 3.2 ∎

### 3.3 Projection Games

In a projection game, the verification predicate is of a special form: to every question $(X, Y)$ posed, there is a function $f_{XY} : B \rightarrow A$ such that if $V(x, y, a, b)$ is true, then $f_{XY}(b) = a$. This is effectively the same as the projection constraints in the Label Cover problem we encountered in Section 2.2.2. Rao [Rao08] proved a stronger version of the parallel repetition theorem for the special case of projection games, that we shall now discuss.

**Theorem 3.11.** (Parallel Repetition for Projection Games, [Rao08]) There is a universal constant $\alpha > 0$ such that if $\mathcal{G}$ is a projection game with $\text{val}(\mathcal{G}) \leq 1 - \epsilon$, then $\text{val}(\mathcal{G}^{\otimes n}) \leq (1 - \epsilon/2)^{\alpha n}$.

There are two ways this result is stronger than the one for general games: firstly the dependence on the answer-set size $c$ is eliminated, and secondly, it decreases faster with $\epsilon$ than in the general case. The proof is essentially the same as the case of general case in terms of method. The main Lemma for projection games gets changed to:
Lemma 3.12. (Main lemma for projection games) For a \( \beta \geq 0 \), for all \( S \subseteq [n] \) such that \( n - |S| \geq \frac{\log(1/5\beta^2)}{2\beta^2} \), if \( \Pr[W_S] \geq 2^{-\beta^2(n-k)} \) then \( \mathbb{E}_{i \in S}[\Pr[W_i|W_S]] \leq 1 - \epsilon + 75\beta \).

Now to see how the proof of the main theorem proceeds from this, set \( \beta = \epsilon/150 \), and use the induction hypothesis as there is a \( S \subseteq [n] \) with \( |S| = k \) such that \( \Pr[W_S] \leq (1 - \epsilon/2)^k \). Suppose \( \Pr[W_S] \geq (1 - \epsilon/2)^k \geq 2^{-c(n-k)/\text{const}} \) (this happens for \( k \leq \text{const} \cdot n \) = \( t \) (say); since \( -\log(1 - \epsilon/2) = O(\epsilon) \)). Then, there is an \( i \notin S \) with \( \Pr[W_i|W_S] \leq 1 - \epsilon/2 \), using Lemma [3.12]. Including this into this set, we have \( \Pr[W_{S \cup \{i\}}] \leq (1 - \epsilon/2)^{k+1} \). Since the induction can proceed until \( k = t = \text{const} \cdot n \), we will have

\[
\Pr[W] \leq \Pr[W_S] \\
\leq (1 - \epsilon/2)^{\text{const} \cdot n} \\
\leq (1 - \epsilon)^{\text{const} \cdot n}
\]

Also, we note that the application of the lemma required that \( n - k \geq n - t \geq \frac{\log(1/5\beta^2)}{2\beta^2} \) (where \( \beta \) is set as above), giving a lower bound \( n_t \) on \( n \) in terms of \( \epsilon \) for the above statement to hold. However, if the theorem holds for large \( n \), it must hold for all \( n \). This is because, if, for some \( l < n_t \), \( \Pr[W_l] > (1 - \epsilon)^{\text{const} \cdot l} \), it can be used to play the game \( G^\otimes n \) independently on blocks of size \( l \) to get \( \text{val}(G^\otimes n) > ((1 - \epsilon)^{\text{const} \cdot l})^{n/l} = (1 - \epsilon)^{\text{const} \cdot n} \), a contradiction.

We now show how Lemma [3.12] is proved. While most of the machinery from the general case remains the same, we use the following stronger lemma to replace Lemma [3.9] as a starting point:

Lemma 3.13. (Rao08) Let \( P \in \mathcal{P} \), \( M_1 \ldots M_l \in \mathcal{M} \), \( N \in \mathcal{N} \) be random variables such that \( \{PM^N\} = \{P\{M_1\}E\ldots\{M_l\}E\} \) and \( E \) be an event. Further, let \( H \subseteq \text{supp}(N) \times \text{supp}(P) \) (the event \( H \) is \( \Pr[N, P \in H] \)). Then,

\[
\mathbb{E}_{i \in H}[|NPM_i|E] - |NPE\{M_i|P\}|1 \leq \sqrt{\Pr[H|E]\log|N_H^*| + \log \left( 1 - \Pr[H|E] \right)}
\]

where \( N_H^* = \arg \max_{P \in \mathcal{P}} |\text{supp}[N|E \land H \land P = p]| \)

The key point in the analysis will be the definition of \( H \), which will allow us to bypass the answer set size. The following lemma does this; the random variables used here are exactly the ones as defined in Section 3.2.3

Lemma 3.14. Let \( H \subseteq \text{supp}(R) \times \text{supp}(A) \) such that \( \Pr[A = a|R = r] \geq 2^{-h} \), where we will fix \( h \) later. Then the following is true:

1. \( \forall r \text{ supp}\{A|R = r \land H\} \leq 2^h \)
2. \( \Pr[W_S \land H] \geq \Pr[W_S] - 2^{-h} \)
3. \( \Pr[H|W_S] \geq 1 - 2^{-h}/\Pr[W_S] \).

Proof. (1) trivially follows from the definition of \( H \).
(2) For every \( r \), \( \{A, B|R = r\} = \{A|R = r\}\{B|R = r\} \). This yields:

\[
\Pr[W_S \land H^C] = \sum_{(f(b), r) \notin H} \Pr[r = r, B = b]\Pr[A = f_Q(b)|R = r] \\
\leq 2^{-h}
\]

Thus, \( \Pr[W_S \land H] \geq \Pr[W_S] - 2^{-h} \)
(3) \( \Pr[H|W_S] = \frac{\Pr[H \land W_S]}{\Pr[W_S]} \geq 1 - 2^{-h}/\Pr[W_S] \)
We now prove the main lemma. Using the assumptions \( \Pr[W_S] \geq 2^{-\beta^2(n-k)} \), and setting \( h = 3\beta^2(n-k) \), after some arithmetic, we get the bounds:

\[
\begin{align*}
\Pr[W_S \wedge H] & \geq 2^{-2\beta^2(n-k)} \\
\Pr[H|W_S] & \geq 1 - 2^{-2\beta^2(n-k)} \\
& \geq 1 - 5\beta^2 \text{ using the bound on } n - k
\end{align*}
\]

Substituting these values into Lemma 3.13, with the instantiations \( P \leftarrow R, M_i \leftarrow U_i, N \leftarrow A \) and \( E \leftarrow W_S, H \) as in Lemma 3.14 we get

\[
\mathbb{E}_{g_S}[\|\{ARU_i|W_S\} - \{AR|W_S\}U_i|R\|_1] \leq \sqrt{\Pr[H|E]h + 2\beta^2(n-k) + \frac{5\beta^2(n-k)}{n-k}} + 5\beta^2
\]

The rest of the proof from the general case holds, with \( \delta = 5\beta \), yielding the result of Lemma 3.12.

### 3.4 Concentration Result

One may ask the question, suppose that the players do not want to win all the co-ordinates in \( G \otimes n \), but only want to win, say, \( \gamma \) fraction of them. What can we say about the probability of this event? We can easily bound the expected number of co-ordinates won in \( G \otimes n \). Introduce a variable \( W_i \) which is the indicator variable for a win in co-ordinate \( i \), and let \( W = \sum_{i=1}^{n} W_i \):

\[
\mathbb{E}_{X^nY^n}[W] = \sum_{i=1}^{n} \mathbb{E}[W_i] \leq \text{val}(G) \times n = (1 - \epsilon)n
\]

The strategy of playing every co-ordinate independently achieves this bound. Rao [Rao08] proves a concentration result for this setting:

**Theorem 3.15.** (Concentration result for general games) There is a universal constant \( \alpha \geq 0 \) such that for \( \text{val}(G) \leq 1 - \epsilon \) and \( \delta \leq \epsilon \), the probability that the players win more than a \( 1 - \epsilon + \delta \)-fraction of the co-ordinates in \( G \otimes n \) is bounded by

\[
\Pr[W \geq (1 - \epsilon + \delta)n] \leq \left(1 - \frac{\delta/2}{1 - \epsilon + 3\delta/4}\right)^r
\]

where \( r = \frac{\text{const.} \beta^2 n + \text{log}(1/\epsilon)}{1 - \epsilon + 3\delta/4} \). This can be simplified as (using \( \log(\frac{1}{\gamma}) = O(\gamma) \)), \( \Pr[W \geq (1 - \epsilon + \delta)n] \leq (1 - \delta/2)^n \delta^2 n/c \) for some constant \( \gamma \).

We note that if the players resort to the independent strategy, then we can use Chernoff bounds to get the following:

\[
\Pr_{\text{(indep)}}[W \geq (1 - \epsilon + \delta)n] \leq \exp(-\delta^2 (1 - \epsilon)n)
\]

The above Theorem gives a somewhat looser dependence on \( \delta \), but one that works for all strategies: \( (1 - \delta)^n \delta^2 n/c \approx \exp(-\alpha \delta^2 n/c) \).

Such a concentration result is useful in the experimental verification of quantum entanglement effects, or EPR paradoxes. We illustrate this by discussing one particular game where quantum effects play a role. We don’t give a background of quantum computation here though, and refer the reader to standard texts on the topic.
3.4.1 The CHSH game

Consider a modified version of the 2-Prover 1-Round game, where the players share an entangled state $|\psi\rangle \in \mathcal{L} \otimes \mathcal{N}$, initiated before the game starts. The space $\mathcal{L}$ consists of the qubits that $P1$ can use, and $\mathcal{N}$ consists of the qubits for $P2$. The extra power that $P1$ and $P2$ possess in the quantum case over the classical one is that depending on their input, each player can perform a quantum measurement on his qubits in $|\psi\rangle$, and can answer according to the outcome. The quantum value of the game, $\text{qval}(G)$, is the maximum probability (over the questions, and randomness inherent in the measurements) of $P1$ and $P2$ of winning $G$, when using a quantum strategy. It turns out that this can give the players quite some advantage over classical strategies.

We now describe the CHSH game. The question set is in $\{0,1\}^2$ and the answer set for both $P1$ (named Alice here) and $P2$ (Bob) is $\{0,1\}$. The verification predicate is

$$V(x, y, a, b) = 1 \iff a \oplus b = x \land y$$

By simply enumerating out the deterministic strategies, it is easy to see that $\text{val}(G) = \frac{4}{3}$. What is the quantum value of this game?

Suppose Alice and Bob share the state $|\psi\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}$, where the first qubit belongs to Alice, the second to Bob. Define the states, for any given $\theta \in [0, 2\pi]$

$$|\phi_0(\theta)\rangle = \cos \theta |0\rangle + \sin \theta |1\rangle$$

$$|\phi_1(\theta)\rangle = -\sin \theta |0\rangle + \cos \theta |1\rangle$$

The quantum strategy used by Alice is: if she receives 0, she measures her qubit with respect to the basis :

$$\{|\phi_0(0)\rangle, |\phi_1(0)\rangle\}$$

This corresponds to the observable $A_0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ with outcomes 1, -1. (Alice identifies answer $a$ with outcome $(-1)^a$).

On receiving 1, she measures her qubit with respect to the basis :

$$\{|\phi_0(\pi/4)\rangle, |\phi_1(\pi/4)\rangle\}$$

This is the observable $A_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

Similarly, Bob’s measurement bases on receiving 0 and 1 are: ($B_0, B_1$ are the corresponding observables)

$$0 \rightarrow \{|\phi_0(\pi/8)\rangle, |\phi_1(\pi/8)\rangle\} \implies B_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

$$1 \rightarrow \{|\phi_0(-\pi/8)\rangle, |\phi_1(-\pi/8)\rangle\} \implies B_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ -1 & -1 \end{pmatrix}$$

We have that $\langle \psi | A_r \otimes B_s | \psi \rangle = \text{Pr}[\text{win on } r, s] - \text{Pr}[\text{lose on } r, s]$ for $rs \in \{0,1\}^2 - \{00\}$, since the LHS is the expected value of the product of Alice and Bob’s measurement outcomes, and they win whenever their outcomes are the same (product $=+1$) and lose when they are different (product $=-1$). On the other hand, for the case of $rs = 11$, $\langle \psi | A_r \otimes B_s | \psi \rangle = \text{Pr}[\text{lose on } r, s] - \text{Pr}[\text{win on } r, s]$. (Note that all these probabilites are over the randomness in the measurement outcomes). Thus, we have finally,

$$\text{Pr}[\text{win}] - \text{Pr}[\text{lose}] = \frac{1}{4} \left[ \langle \psi | A_0 \otimes B_0 | \psi \rangle + \langle \psi | A_0 \otimes B_1 | \psi \rangle + \langle \psi | A_1 \otimes B_0 | \psi \rangle - \langle \psi | A_1 \otimes B_1 | \psi \rangle \right]$$

$$\approx \frac{1}{2} \frac{1}{2\sqrt{2}} \approx 0.85$$

$$\implies \text{Pr}[\text{win}] = \frac{1}{2} + \frac{1}{2\sqrt{2}} \approx 0.85$$
Use of Parallel Repetition

Thus we see that for the CHSH game, $\text{val}(\mathcal{G}) = 0.75$, whereas $\text{val}(\mathcal{G}) \geq 0.85$. Repeat this game in parallel $n$ times, and consider if the players win, if they win more than $0.8$-factor of the co-ordinates. Using the Chernoff bound in the quantum case suggests that the probability of the quantum players losing goes down exponentially in $n$. The concentration bound for classical games implies that classical players can win only with exponentially decreasing probabilities. Thus, this (new) game has a very large gap between quantum and classical strategies.

3.4.2 Proof of the concentration result

Pick a random subset $S \subseteq [n]$ of size $t$ (we will fix $t$ later), and find the probability of the players winning on $S$. Consider the method of generating $S$ randomly as a sequence $\{i_1 \ldots i_t\}$ where we call $S_j = \{i_1 \ldots i_j\}$ and $i_{j+1}$ is chosen uniformly at random from $[n] - S_j$. We show the following (recall that $D$ is the distribution over the questions):

Lemma 3.16. For $t = \frac{\gamma^2 n}{\epsilon + \gamma^2 + \log(\frac{1}{\gamma \delta})}$ where $15\gamma = \delta/4$, $\Pr_{S_k, D}[W_{S_k}] \leq 2(1 - \epsilon + \delta/4)^t$

Proof. For any $k \leq t$, there are two options: call $S_k$ good, if for that (fixed) $S_k$, $\Pr_D[W_{S_k}] \geq (1 - \epsilon + \delta/4)^t$ and bad otherwise

$$
\Pr_{S_k, D}[W_{S_k}] \leq \Pr_{S_k, D}[W_{S_k} \wedge S_k \text{ is good}] + \Pr_{S_k, D}[W_{S_k} \wedge S_k \text{ is bad}]
\leq \Pr_{S_k, D}[W_{S_k} \wedge S_k \text{ is good}] + \Pr_{S_k, D}[W_{S_k} | S_k \text{ is bad}]
\leq \Pr_{S_k, D}[W_{S_k} \wedge S_k \text{ is good}] + (1 - \epsilon + \delta/4)^t
$$

We show that $\Pr_{S_k, D}[W_{S_k} \wedge S_k \text{ is good}] \leq (1 - \epsilon + \delta/4)^t$. Firstly, observe that $\Pr_D[W_{S_k}] \geq (1 - \epsilon + \delta/4)^t \implies \Pr_D[W_{S_{k-1}}] \geq (1 - \epsilon + \delta/4)^t$. Further by the setting of the parameter $t$,

$$(1 - \epsilon + 15\gamma)^t \geq 2^{-\gamma^2(n-t)+tc} \geq 2^{-\gamma^2(n-k)+kc} \forall k \leq t$$

Then we can apply the main lemma for general games Lemma 3.2 to conclude that

$$
\Pr_{S_k, D}[W_{S_k} | W_{S_{k-1}}] \leq (1 - \epsilon + \delta/4)
\implies \Pr_{S_k, D}[W_{S_k} | W_{S_{k-1}} \wedge S_{k-1} \text{ good}] \leq (1 - \epsilon + \delta/4) \quad \text{we started out with this, so conditioning is ok}
\implies \Pr_{S_k, D}[W_{S_k} | W_{S_{k-1}} \wedge S_{k-1} \text{ good}] \leq (1 - \epsilon + \delta/4) \quad \text{since it holds for every good } S_{k-1}
$$

Hence, we have:

$$
\Pr_{S_k, D}[W_{S_k} \wedge S_k \text{ is good}] = \prod_{j=1}^{t} \Pr_t[W_{S_j} \wedge S_j \text{ is good}|W_{S_{j-1}} \wedge S_{j-1} \text{ is good}] 
\leq (1 - \epsilon + \delta/4)^t
$$

Let $Z$ denote the number of games won. Whenever $Z \geq (1 - \epsilon + \delta)n$ co-ordinates are won, pick a random subset $S$ of size $t$ from the set where the players won, and blame it for the win. Then we have:
\[ \Pr[Z \geq (1 - \epsilon + \delta)n] \leq \Pr[\text{some } S \text{ is blamed}] \]
\[ \leq \sum_S \Pr[S \text{ is blamed}] \quad \text{(union bound)} \]
\[ \leq \sum_S \Pr[W_S \mid \Pr[S \text{ is chosen}\mid \text{we win in } \geq (1 - \epsilon + \delta)n \text{ coordinates}] \]
\[ \leq \left( \frac{n}{t} \right) (1 - \epsilon + \delta/4)^t \left( \frac{(1 - \epsilon + \delta)n}{t} \right)^{-1} \]
\[ \leq \left( \frac{n(1 - \epsilon + \delta/4)}{n(1 - \epsilon + \delta) - t} \right)^t \quad \text{(expanding out factorials) (3.7)} \]
\[ \leq \left( 1 - \epsilon + \delta/4 \right)^t \left( \frac{1 - \epsilon + 3\delta/4}{1 - \epsilon + \delta} \right)^t \quad \text{(using } t/n \leq \delta/4) \]

This concludes the proof of the concentration result. \( \square \)

We can exactly follow the same steps for the concentration bound, by using the main lemma for projection games, Lemma \( 3.12 \) and setting \( t \) to be such that:
\[ (1 - \epsilon + 75\beta)^t = \frac{2 - \beta}{\beta^2} \]
\[ \Rightarrow \frac{t}{n} = \frac{\beta^2}{\beta^2 + \log \left( \frac{1}{1 + 75\beta} \right)} \]

The only point of concern here is that \( \frac{t}{n} \) cannot be replaced with \( \frac{\delta}{4} \) in the denominator in Eqn. (3.7).

Set \( 75\beta = \frac{\delta}{4} \), and let \( \alpha \triangleq t/n = \frac{1}{1 + (\frac{300^2}{\delta^2}) \log \left( \frac{1}{1 - \epsilon + 3\delta/4} \right)} \). Following the steps of the previous proof, we get that
\[ \Pr[Z \geq (1 - \epsilon + \delta)n] \leq \left( \frac{1 - \epsilon + \delta/4}{1 - \epsilon + \delta - \alpha} \right)^t \]

Note that for our setting of parameters, we will have that \( 1 - \epsilon + \delta - \alpha > 0 \). We will show that there is a constant \( \gamma \) such that \( \left( \frac{1 - \epsilon + \delta/4}{1 - \epsilon + \delta - \alpha} \right)^t \leq 1 - \delta \gamma \). To show this, it is sufficient to set \( \gamma \) such that:
\[ \delta \gamma \leq \frac{3\delta}{4} - \alpha \]
\[ \Leftrightarrow \gamma \leq \frac{3}{4} - \frac{\alpha}{\delta} \]
\[ \Leftrightarrow \frac{\epsilon}{\delta} \leq \frac{3}{4} - \frac{300^2}{\delta} \log \left( \frac{1}{1 - \epsilon + 3\delta/4} \right) \]
\[ \Leftrightarrow \frac{\epsilon}{\delta} \leq \frac{3}{4} - \frac{300^2}{\delta^2} \log \left( \frac{1}{1 - \epsilon + 3\delta/4} \right) \]
\[ \Leftrightarrow \gamma \leq \frac{3}{4} - \frac{1}{300^2} \left( \text{since } \frac{\epsilon}{\log \left( \frac{1}{1 - \epsilon + 3\delta/4} \right)} \right) \]

Now, exactly the same proof goes through, provided \( n \geq t + \log(1/5\beta^2) \), call this value \( n_0 \). However, as before, if the result holds for large \( n \), it must hold for all \( n \): suppose the players win on some subset of size \( l \) (where win is defined as winning more than \( (1 - \epsilon + \delta) \)-fraction of the co-ordinates) with probability greater than \( (1 - \gamma \delta)^{l/(\delta l)} = p^l \) (say). Then, for large \( n \geq n_0 \), they could simply apply this strategy independently on each block of size and win a total of greater than \( (1 - \epsilon + \delta) \) fraction of the co-ordinates with probability greater \( (p^l)^n/l = p^n \). This is a contradiction. Thus, we get that:
\[ \Pr[Z > (1 - \epsilon + \delta)n] \leq (1 - \gamma \delta)^{\text{const.} \delta^2 n} \]

\( \square \)
4 Raz’s Counterexample and Unique Games

In the previous section, we saw how the probability of winning the $n$-repeated game goes down exponentially in $n$. However, the concretely known lower bound (i.e. strategy) for any non-trivial game was simply one of playing every co-ordinate independently; this gives $\mathrm{val}(\mathcal{G}^\otimes n) \geq (1 - \epsilon)^n$ (where $\mathrm{val}(\mathcal{G}) = (1 - \epsilon)$).

The obvious question to ask is, is the dependence on $\epsilon$ in the Parallel Repetition theorem tight? The following conjecture is usually called Strong Parallel Repetition:

**Conjecture 4.1.** (Strong Parallel Repetition) For any game $\mathcal{G}$ with $\mathrm{val}(\mathcal{G}) = (1 - \epsilon)$, $\mathrm{val}(\mathcal{G}^\otimes n) \leq (1 - \epsilon)^\Omega_c(n)$ (where the constant in the exponent may depend upon game parameters, such as answer set size).

As it turns out, a Strong Parallel Repetition theorem for general games (or even projection games) would imply certain inapproximability results based on the Unique Games Conjecture (UGC). However, a counterexample to the Strong Parallel Repetition conjecture was recently discovered by Raz, which we turn to immediately. Later in this section, we look at the UGC and the role of parallel repetition therein.

4.1 Counterexample to Strong Parallel Repetition

In [Raz08], Raz gave an explicit example of a game that has $\mathrm{val}(\mathcal{G}) = (1 - \epsilon)$, but $\mathrm{val}(\mathcal{G}^\otimes n) \geq (1 - \epsilon^2)^\Omega(n)$. We describe the game below:

4.1.1 MAX-CUT on an odd cycle : $\mathcal{G}_{C_m}$

The game is played on a graph $G = (V, E)$ that is a cycle of length $m$, $m$ is odd. For now, number the vertices on the graph from $0 \ldots m - 1$. The game proceeds as follows:

- The verifier picks a vertex $v \in \{0, \ldots, m - 1\}$ uniformly at random and sends it to $P1$
- With probability $\frac{1}{2}$, $P2$ gets $i$, and with probability $\frac{1}{2}$, he gets one of the neighbors of $i$ (that is, gets $i + 1$ (mod $m$) w.p. $\frac{1}{4}$ and $i - 1$ (mod $m$) w.p. $\frac{1}{4}$).
- The provers respond with a bit $\in \{0, 1\}$ that indicates the color of their vertex
- The predicate is $V(x, y, a, b) = 1 \iff (x == y \land a == b) \lor (x \neq y \land a \neq b)$
  - This means that the provers must assign consistent colours to the same vertex, and must not colour an edge monochromatically

![Figure 4.1: Odd Cycle game](image)

*Value of the game*
Suppose the provers fix a colouring strategy for the vertices beforehand, and play according to that (as in Figure 4.1b). Since the cycle is odd, one edge is monochromatic, and the probability of the provers getting this is $\frac{1}{2m}$. They get all other answers right, so we have $\text{val}(G_{C_m}) \geq 1 - \frac{1}{2m}$.

On the other hand, for any other strategy, let $P_1$ have $f_1 : V \to \{0, 1\}$ and $P_2$ have $f_2 : V \to \{0, 1\}$. Then, if $\exists v \in V : f_1(v) \neq f_2(v)$ then they err with probability $\frac{1}{2m}$ (when they both receive $v$). On the other hand, if $f_1(v) = f_2(v) \forall v \in V$, then this is simply the previous case, and they err on at least one edge with probability $\frac{1}{2m}$. Thus,

$$\text{val}(G_{C_m}) = 1 - \frac{1}{2m}$$

**Remark 4.2.** As in the CHSH game, a quantum strategy succeeds with higher probability; $\text{qval}(G_{C_m}) \geq \cos^2 \frac{\pi}{4} m \geq 1 - \left(\frac{\pi}{4m}\right)^2$ (see [CHTW04]).

**Theorem 4.3.** (Raz, [Raz08]) $\text{val}(G_{\otimes n C_m}) \geq 1 - O\left(\frac{\sqrt{n}}{m}\right)$.

Just to gauge the effect of the above theorem in terms of the notation we have used until now, let $\frac{1}{2m} = \epsilon$. Then $\text{val}(G_{C_m}) = 1 - \epsilon$, and $\text{val}(G_{\otimes n C_m}) \geq 1 - \epsilon \sqrt{n}$, where $\epsilon$ is a constant. For $n_0 = 1/2e^2\epsilon^2$ rounds, the value of the above game is higher than $\frac{1}{2}$. So, for $n$ larger than $n_0$, we simply divide $n$ into blocks of size $n_0$, and play the game independently on those blocks. Thus, the value of the game will then satisfy $\text{val}(G_{\otimes n C_m}) \geq 2^{O(\epsilon^2 n)} \approx (1 - \epsilon^2)^{O(n)}$. This shows that Rao’s bound for projection games is essentially tight, in terms of dependence on $\epsilon$, since $G_{C_m}$ is a projection game.

**Remark 4.4.** The above game strategy is said to survive $O\left(\frac{1}{\epsilon^2}\right)$ rounds. The independent strategy survives around $O\left(\frac{\epsilon^2}{2}\right)$ rounds, since $(1 - \epsilon)^{O(\frac{1}{\epsilon})} \approx \exp(-\text{const})$.

### 4.1.2 Proof of the Theorem

To prove 4.3 we need to give a strategy for $P_1$ and $P_2$. The strategy will be probabilistic (using shared randomness), and interestingly, uses the correlated sampling method used in Section 3.2.2. To recap, we state the method in slightly different language here:

**Theorem 4.5.**(Correlated sampling, Restated) Consider two probability distributions $D_X, D_Y$ over the same domain $\Omega$, such that $\|D_X - D_Y\|_1 \leq \epsilon$. If two non-communicating players $P_1$ and $P_2$ each have knowledge of exactly one of these distributions; $P_1$ knows $D_X$ and $P_2$ knows $D_Y$, then they can use shared randomness $R$ over some domain $\mathcal{R}$ to produce samples from $\Omega$ such that $P_1$ draws $X \sim D_X$, $P_2$ draws $Y \sim D_Y$, and $\text{Pr}[X \neq Y] \leq 2\epsilon$. The space $\mathcal{R}$ is independent of $D_X$ and $D_Y$. (See Fig. 2)
Remark 4.6. Note that if they were allowed to communicate, there is a shared space they can draw from such that the marginals are respected, and $Pr[X \neq Y] \leq \epsilon$. In this case, $R$ depends upon $D_X$ and $D_Y$.

The 1-round game:

To warm up for the $n$-round game, let us consider a trivial correlated sampling strategy for the 1-round game $G_{C_m}$: the players do not fix a single colouring to answer from, but rather draw the colouring strategy uniformly at random from the set of (effectively) $m$ possible colouring strategies. Every colouring strategy uniquely corresponds to the one edge that is monochromatic under that colouring; so effectively, they are drawing a single edge as advice, using which they infer the colours of their received vertices.

Here, the distribution over the colouring strategies used is the same for both players, so they always draw the same strategy. The probability that they err is simply the probability that (a) they receive adjacent vertices ($Pr = 1/2$), and (b) the drawn edge corresponds to the one received ($Pr = 1/m$). Thus, they lose with probability $\frac{1}{2m}$, and as we have seen, they can’t do any better.

The $n$-round game:

In the $n$-round game, $P1$ gets $X^n = (X_1, \ldots, X_n)$ and $Y^n = (Y_1, \ldots, Y_n)$. Let $m = 2k + 1$, and $K \equiv \{-k, \ldots, +k\}$. As in the one-round strategy, they will draw colouring advice from the shared source. However, there are two main differences here: firstly, instead of simply drawing one edge, they are drawing a single edge as advice, using which they infer the colours of their received vertices. They note that at the domain edges, the edges are labelled by the number of the vertex they are opposite to. With this labelling on the vertices, the edges are labelled by the number of the vertex they are opposite to. For $e \in E$, let $l_v(e)$ be the numbering assigned to edge $e$ with respect to $v$ as the reference vertex.

Define the probability density $f(\cdot)$ over the domain $K$, as follows (here $\gamma$ is a normalization factor): $$f(i) = \gamma (k - |i| + 1)^2$$

We note that at the domain edges, $f(k), f(-k) \leq O\left(\frac{1}{m^2}\right)$.

For convenience, given a reference vertex $v$, we label the vertices in $C_m$ (the cycle on $m$ vertices) using $K$ with $v$ labelled as 0, and $u \leftarrow d(v \rightarrow u)$, i.e. the distance of moving from $v \rightarrow u$, considering anticlockwise direction as negative.

With this labelling on the vertices, the edges are labelled by the number of the vertex they are opposite to. For $e \in E$, let $l_v(e)$ be the numbering assigned to edge $e$ with respect to $v$ as the reference vertex.

Define the probability distribution induced over the edges $E$ with respect to vertex $v$ as: $$P_v(e) = f(l_v(e))$$

In particular, we note that the probability of drawing an edge that touches $v$ is at most $O\left(\frac{1}{m^2}\right)$.

Suppose $P1$ receives vertex $X^n$ and $P2$ receives vertex $Y^n$. The distribution $P1$ will look to draw advice from (this is over $E^n$) is: $$D_{X^n}(\vec{e}) = \prod_{i=1}^{n} P_{X_i}(e_i)$$

Similarly, $P2$ looks to draw advice from: $$D_{Y^n}(\vec{e}) = \prod_{i=1}^{n} P_{Y_i}(e_i)$$

If these distributions are close enough, then they could use correlated sampling to draw the same advice on all $n$ co-ordinates. Suppose that $E_{X^n,Y^n}\|D_{X^n} - D_{Y^n}\|_1 \leq \delta$. Then the probability of them drawing the same strategy on all co-ordinates is $\geq 1 - 2\delta$. Once they have drawn the same strategy, say $\vec{e} = (e_1, \ldots, e_n)$, they err if $\exists i : e_i = \{X_i, Y_i\}$. However, then

$$Pr[\exists i : e_i = \{X_i, Y_i\}] \leq \sum_{i \in [n]} Pr[e_i = \{X_i, Y_i\}] \quad \text{(union bound)}$$

$$\leq n Pr[e_1 \text{ touches } X_1]$$

$$\leq O\left(\frac{n}{m}\right)$$
Then the total error can be bounded as:

\[
\Pr[\text{lose}] \leq 2\delta + (1 - 2\delta)O\left(\frac{n}{m^3}\right) \tag{4.1}
\]

What we have to now do is bound the quantity \(\delta\) for the distributions we have. We do this using the notion of Hellinger Distance between distributions.

### 4.1.3 Hellinger distances, finishing the proof

The Hellinger distance between two probability densities \(P, Q\) over a domain \(\Omega\) is defined as:

\[
H^2(P, Q) = \frac{1}{2} \int_{\Omega} (\sqrt{P} - \sqrt{Q})^2
\]

Over a discrete domain, this is really \(\frac{1}{2} \sum_{\omega \in \Omega} \left(\sqrt{P(\omega)} - \sqrt{Q(\omega)}\right)^2\). If \(\Omega\) is finite of cardinality \(d\), then we can view \(P, Q\) as vectors in \(\mathbb{R}^d\). Let \(u = \sqrt{P}, v = \sqrt{Q}\) be vectors in \(\mathbb{R}^d\) obtained by taking the squareroot of each co-ordinate of the distributions. Then the Hellinger distance is basically \(\frac{1}{2}\|u - v\|^2_1\), where \(u, v\) are unit norm vectors in \(\mathbb{R}\). We state some properties of Hellinger distances here that we will use, and relegate their proofs to the Appendix.

1. \(H^2(P, Q) \in [0, 1]\)
2. \(H^2(P, Q) \leq \|P - Q\|_1 \leq \sqrt{2}H(P, Q)\)
3. \(H^2(P_1 \otimes P_2 \otimes \ldots \otimes P_l, Q_1 \otimes Q_2 \otimes \ldots \otimes Q_l) \leq \sum_{i=1}^l H^2(P_i, Q_i), \) where \(P_1 \otimes \ldots \otimes P_l\) is the product distribution over the domain \(\Omega_1 \otimes \ldots \otimes \Omega_l\), similarly for \(Q_1 \otimes \ldots \otimes Q_l\)

**Claim 4.7.** \(\|P_1 \otimes \ldots \otimes P_l - Q_1 \otimes \ldots \otimes Q_l\|_1 \leq \sqrt{2^n} \sum_{i=1}^l H^2(P_i, Q_i)\)

**Proof.** Immediate from points 2, 3 above. \(\square\)

### Final Part of the Proof:

Returning back to the proof, we observe that the distributions that the players use \(D_{X^n}, D_{Y^n}\) can be viewed as product distributions over the domain \(K^n\). The first co-ordinate of the domain \(K_1\), corresponds to a labelling of the edges with respect to reference \(X_1\). Thus, \(D_{X^n}(i) = \gamma(k + 1 - |i|)^2\). Since \(Y_1\) is either adjacent to \(X_1\), or is \(X_1\) itself, \(D_{Y^n}\) is either the same distribution as \(D_{X^n}\), or it is a slightly shifted one: if \(Y_1\) is labelled as 1 with reference to \(X_1\), then we will have:

\[
D_{Y_1}(i) = \begin{cases} 
\gamma(k + 1 - |i - 1|)^2 & i \neq -k \\
\gamma & i = -k
\end{cases}
\]

and similarly for \(Y_1\) labelled as \(-1\). We now look to use Claim 4.7 to conclude:

\[
\|D_{X^n} - D_{Y^n}\|_1 \leq \sqrt{2} \sum_{i \in [n]} H^2(D_{X_i}, D_{Y_i})
\]

But, every term in the sum can be bounded, since even in the worst case of \(X_i, Y_i\) being adjacent, we will have:

\[
H^2(D_{X_i}, D_{Y_i}) \leq \frac{\gamma}{2} \left( \sum_{|j - k| \in K - \{-k\}} (k + 1 - |j|) - (k + 1 - |j - 1|) \right) + O\left(\frac{1}{m^2}\right)
\]
This gives us:
\[ \|D_X - D_Y\|_1 \leq O(\sqrt{\frac{n}{m}}) \]
Thus, combining with where we left off by substituting this value for \( \delta \), we get
\[ \Pr[\text{Win}] \geq 1 - O(\sqrt{\frac{n}{m}}) - O\left(\frac{n}{m^3}\right) \]
Since we can always play the game in blocks of \( n \leq \alpha m^2 \) for a small enough constant \( \alpha \), and repeat this (in fact due to the term \( \sqrt{n/m} \), this is actually necessary), the last term is also \( O(\sqrt{n/m}) \). So overall, we have:
\[ \text{val}(G^{\otimes m}) \geq 1 - O\left(\frac{\sqrt{n}}{m}\right) \]

4.2 QP and SDP formulations

Feige and Lovasz [FL92] give a quadratic programming formulation to determine the value of a game. This can then be relaxed to a semidefinite program, which yields results regarding the parallel repetition of some classes of games.

4.2.1 QP formulation

We can form a natural quadratic program to find the value of any game \( G \). Encode each player’s strategy as a binary vector: \( P_1 \)’s binary vector \( p \) has a length of \( |X||A| \), with \( p_{xa} = 1 \iff f_1(x) = a \) and \( P_2 \)’s vector has a length of \( |Y||B| \), with a similar encoding. Consider the following quadratic program, where the cost matrix \( C \) is of dimensions \( |X||A| \times |Y||B| \), and has each entry as:
\[ \{C_{xyab}\} = \Pr[X = x, Y = y]V(x, y, a, b) \]

\[
\max_{p,q} \quad p^T C q \\
\text{subject to} \quad \sum_{a \in A} p_{xa} = 1 \quad \forall x \in X \quad (4.2) \\
\sum_{b \in B} q_{yb} = 1 \quad \forall y \in Y \quad (4.3) \\
p \geq 0, q \geq 0
\]
Constraints (4.2), (4.3) ensure that each player can give only one answer per question. Denote the maximum of the above program as \( \sigma(G) \).

**Theorem 4.8.** ([FL92]) \( \text{val}(G) = \sigma(G) \)

**Proof.** It is obvious that \( \text{val}(G) \leq \sigma(G) \), since a valid strategy for the provers is a feasible solution to the above program. However, since the solutions may have fractional components, the other way requires an argument. Since the constraints are all linear, we observe that for any fixed value of \( p \), the optimum for \( q \) is obtained at an extreme point (since the program becomes linear in \( q \)), which is a \( 0 - 1 \) vector. Similarly, for any fixed value of \( q \), the optimum is obtained at a \( 0 - 1 \) vector for \( p \). These taken together imply that the optimal value will be attained at some binary \( p, q \). These can be decoded back to unique solution strategies for the provers, who will win the game with probability \( \sigma(G) \). \( \square \)
4.2.2 SDP formulation

We can relax the QP in the preceding section to yield a SDP. Define the game \( G_1 \otimes G_2 \) in the natural way: a two co-ordinate game, with one for each game. The distribution over the questions is also the product distribution of the questions of individual games. For \( G_1 = G_2 \), this becomes \( G_1^{\otimes 2} \).

One reason we would like relax the QP further is that the value of the QP is supermultiplicative under parallel repetition: \( \sigma(G_1 \otimes G_2) \geq \sigma(G_1) \sigma(G_2) \) (this is to be expected, in the light of Theorem 4.8). A relaxation that follows submultiplicativity under repetition would be useful for proving upper bounds.

Relaxation of the QP is done in the usual way: replace the scalars that follows submultiplicativity under repetition would be useful for proving upper bounds.

The following theorem asserts that this is actually an equality:

**Theorem 4.9.** \( \omega(G_1 \otimes G_2) = \omega(G_1) \times \omega(G_2) \)

Relaxation of the QP is done in the usual way: replace the scalars that follows submultiplicativity under repetition would be useful for proving upper bounds.

**Corollary 4.10.** \( val(G^{\otimes n}) \leq (\omega(G))^n \)

4.2.3 SDPs and Parallel Repetition of Unique games

For a special class of games, called Unique Games, the value of the SDP can be used to characterize the value of the repeated game. As we will mention later, SDPs have played a central role in actually determining the value of 2-Prover 1-Round games and have, of late, been a key method in obtaining hardness of approximation results. We first define a unique game:

**Definition 4.11.** (Unique game) A game \( G \) is said to be unique if:

- \( |A| = |B| \), we call this quantity the **alphabet size** of the game
- for every question pair \( x \in X, y \in Y \), there is a bijection \( \pi_{xy} : A \rightarrow B \) such that \( V(x, y, a, b) = 1 \iff \pi_{xy}(a) = b \)

\(^1\)Although the SDP considered here is slightly different from theirs, which had some more constraints involving cross terms, the formulation is essentially the same
A unique game is a special case of a projection game. Further, a 2-Prover 1-Round unique game can be viewed as a special case of a Label Cover instance, where the constraints on the edges are bijections.

The SDP in the previous section can be simplified for the case of unique games. We can identify the answer sets of both players, so we will assume here that $\mathcal{A} = \mathcal{B} = \{1, \ldots, k\}$, and modify the permutations accordingly. Since we will be working with vectors, to aid notation, we will denote a question from the players err with probability

$$\omega(G) \equiv \max_{\{u_{i}\}, \{v_{i}\}} \mathbb{E}_{(u,v)} \sum_{i \in [k]} \langle u_{i}, v_{x_{u_{i}}(i)} \rangle$$

subject to

\begin{align}
\sum_{i \in [k]} \|u_{i}\|^{2} &= 1 \quad \forall u \in \mathcal{X} \\
\sum_{i \in [k]} \|v_{i}\|^{2} &= 1 \quad \forall v \in \mathcal{Y} \\
\langle u_{i}, v_{i'} \rangle &= 0, \langle v_{i}, v_{i'} \rangle = 0 \quad \forall i, i' \in [k] : i \neq i' 
\end{align}

(4.7) (4.8) (4.9)

Further, we may fold in the question set to $\mathcal{V} = \mathcal{X} \cup \mathcal{Y}$, with the distribution ensuring that questions drawn are “bipartite”. Thus, we can write:

$$\omega(G) \equiv \max_{\{u_{i}\}, \{v_{i}\}} \mathbb{E}_{(u,v)} \sum_{i \in [k]} \langle u_{i}, v_{x_{u_{i}}(i)} \rangle$$

subject to

\begin{align}
\sum_{i \in [k]} \|u_{i}\|^{2} &= 1 \quad \forall u \in \mathcal{V} \\
\langle u_{i}, u_{i'} \rangle &= 0 \quad \forall i, i' \in [k] : i \neq i', \forall u \in \mathcal{V} 
\end{align}

(4.10) (4.11) (4.12)

Observe that the objective above can be rewritten as $\frac{1}{2} \mathbb{E}_{uv} \sum_{i \in [k]} \|u_{i} - v_{x_{u_{i}}(i)}\|^{2}$. We will denote the SDP objective value (using the above program) for a unique game $G$ by $\omega(G)$.

Barak et. al. [BHH+08] show the following result for the $n$-repeated unique game:

**Theorem 4.12.** (Parallel Repetition of Unique Games [BHH+08]) If $G$ is a unique game with alphabet size $k$, and $\omega(G) \geq 1 - \delta$, then

$$\text{val}(G^{\otimes n}) \geq 1 - O(\sqrt{n \delta \log(k/\delta)})$$

Subsequently, Steurer [Ste10] has improved the result to remove the $\log(\frac{1}{\delta})$ factor, yielding $\text{val}(G^{\otimes n}) \geq 1 - O(\sqrt{n \delta \log k})$. We see that the strategy this suggests survives $O(\frac{1}{\delta} \log k)$ repetitions giving that

$$\text{val}(G^{\otimes n}) \geq (1 - \delta)^{\Omega(n \log k)}$$

Combining this with 4.10 we get that $\omega(G)^{\Omega(n \log k)} \leq \text{val}(G^{\otimes n}) \leq \omega(G)^{n}$, implying that the value of the unique game under parallel repetition depends on the SDP value. In particular, games that fool the SDP, i.e. have a large integrality gap, would be expected to violate strong parallel repetition.

We leave out the proof of Theorem 4.12. The underlying idea is the following: given an SDP solution to $G$, we would like to extract out prover strategies from them. This is usually done by using random projections of the solution vectors, and rounding the result. In this case, the provers use shared randomness to sample solution strategies for the question they receive; if $P1$ gets question $u$ and $P2$ gets question $v$, they sample a high dimensional vector $r$ from a random space to determine their answers. Here $P1$ samples using distribution $D_u$, and $P2$ samples using the distribution $D_v$, where $D_u, D_v$ are distributions determined by the SDP solutions. The key point in the proof involves showing that the $H^2(D_u, D_v)$ (in average over $(u, v)$) is bounded by roughly $1 - \sigma(G)$, i.e. $O(\delta)$, implying that the players err with probability $O(\sqrt{\delta})$. Like in Raz’s counterexample, the strategy for the $n$-repeated game is then to simply sample advice from the product distribution: $D_{uv} = \prod_{i \in [n]} D_{u_{i}}$. The Hellinger distance property then ensures that the players err with probability $O(\sqrt{n \delta})$ (ignoring logarithmic factors in the alphabet size). In some sense, the player’s strategies can be viewed as a generalization of the strategy used on $G^{\otimes n}$ to the class of unique games.
4.3 The Unique Games conjecture

Recall that we had looked at Label Cover as a starting point for proving hardness of approximation results, using from Theorem 2.15. However, are there specific classes of Label Cover that are also hard? Consider the subclass of label cover instances consisting of unique games; i.e. as described above, the constraints on every edge are permutations. The following conjecture by Khot makes such a claim about this class.

Denote by $UG_k$ the class of Unique Game problems over the alphabet $[k]$ (this is without loss of generality, as all vertices are assigned labels from a set of the same size, here $k$). An instance of this class will be denoted by $U$, and will have parameters $G(L, R, E), [k], \{\pi_e\}_{e \in E}$, where $G(L, R, E)$ is the (bipartite, with vertex sets $L$ and $R$) constraint graph corresponding to $U$ and $\{\pi_e\}_{e \in E}$ the set of bijective constraints, one on every edge. Note that a $UG_k$ instance defined this way is equivalent to a 2-Prover 1-Round unique game. As before, we will call $OPT(U)$, or the value of $U$ to be the maximum fraction of constraints(edges) in $U$ satisfiable by any labelling: $L \cup R \rightarrow [k]$.

Conjecture 4.13. (Unique Games Conjecture, UGC [Kho02]) For any $\epsilon, \delta \geq 0$, there is a large enough $k = k(\epsilon, \delta)$ such that it is NP-Hard to decide which of the following classes a given $U \in UG_k$ belongs to:

- $YES = \{U \in UG_k : OPT(U) \geq 1 - \epsilon\}$
- $NO = \{U \in UG_k : OPT(U) \leq \delta\}$

(We do not care about $U$ not in either of these classes)

Following the notation used in the Label Cover section, we will call the above (promise) decision problem as $\text{gap}_{1-\epsilon, \delta} - UG_k$. Unlike the problem of $\text{gap}_{1-\epsilon} - LC_\Sigma$, it is easy to decide if a unique game has value 1 or not: fixing a label on a vertex determines the labels on all vertices reachable from this vertex. Thus, cycling through the possible labels in one vertex of a connected component, we can find out if the component is fully satisfiable or not; repeating this for every component gives us a perfect labelling, if one exists.

Further, we note that in order for the conjecture to hold, $k \geq \frac{1}{\epsilon}$. This is because a random labelling of the vertices will satisfy a $\frac{1}{\epsilon}$ fraction of edges in expectation; so every UG instance is easily $\frac{1}{\epsilon}$-satisfiable. Often, in what follows, we will ignore the fact that the constraint graph is bipartite, which effectively means we are working with a weaker form of the above conjecture.

4.3.1 UG-Hardness

In order to show that a (maximization) problem class $\text{gap}_{1-\epsilon} - \mathcal{P}$ having $YES$ and $NO$ instances is hard to decide assuming the UGC, one uses the usual method of producing a polynomial-time gap-preserving reduction from $\text{gap}_{1-\epsilon, \delta} - UG_k$ to $\text{gap}_{1-\epsilon} - \mathcal{P}$. The reduction $\phi$ will satisfy that $U \in YES \implies OPT(\phi(U)) \geq \epsilon$ and $U \in NO \implies OPT(\phi(U)) \leq \delta$.

Various hardness results have been shown assuming the UGC. As an illustrative example, consider the MAX-CUT problem we have encountered before in Section 2.1.1. Using methods similar to the previous section, one can produce an SDP relaxation for a MAX-CUT instance $G(V, E)$:

$$\max_{\{v_i\} \in V} \frac{1}{|E|} \sum_{(i, j) \in E} \frac{1 - \langle v_i, v_j \rangle}{2}$$

(4.13)

$$\forall i \in V : \|v_i\| = 1$$

(4.14)

Let $OPT(G)$ denote the actual value of the maximum cut. Goemans and Williamson give a randomized rounding algorithm, that extracts a cut from the solution vectors of the SDP, while giving a guarantee that the produced solution $A(G)$ satisfies $|A(G)| \geq \alpha_{GW} OPT(G)$, in expected sense (over the randomness in the rounding algorithm). Here, $\alpha_{GW}$ is a constant $\approx 0.875$.

A result by Khot et. al. [KMO04] shows a reduction from UG instances to MAX-CUT instances, thereby concluding that assuming the UGC, it is NP-hard to approximate MAX-CUT to a factor better than $\alpha_{GW}$! We say that MAX-CUT is UG-hard to approximate to a factor better than $\alpha_{GW}$. 

34
In fact, Raghavendra [Rag08] shows that for any CSP, assuming that the UGC is true, the best known approximation algorithm can be found using a canonical SDP relaxation, followed by a rounding procedure on the SDP solutions. The SDP above is essentially the specific case of the canonical one for MAX-CUT. Besides MAX-CUT, other similarly tight inapproximability results have been shown assuming the UGC: for instance, Vertex Cover is UG-hard to approximate to within $2 - \epsilon$, etc. See Kho10 for an overview of recent work on the UGC.

### 4.3.2 Parallel Repetition and UGC

Given that the UGC implies that many well-known problems are NP-hard to approximate to within a given factor, one might ask whether the inapproximability of some of these problems is, in fact, equivalent to the UGC. That is, are there any “UG-complete” problems? The following connection was pointed out by Feige et.al. [FKO07]. Consider the problem of MAXCUT, and suppose we start off with the following conjecture:

**Conjecture 4.14.** (MAX-CUT conjecture) There is a constant $\beta > 0$ such that for every $\delta > 0$, the following problem is NP-hard: Given a MAX-CUT instance graph $G$, distinguish between the two cases:

- **YES** = $\{G : \text{OPT}(G) \geq 1 - \delta\}$
- **NO** = $\{G : \text{OPT}(G) \leq 1 - \beta \sqrt{\delta}\}$

That is, gap$_{1 - \delta, 1 - \beta \sqrt{\delta}}$ - MAXCUT is NP-hard.

We note that the above conjectured hardness is essentially tight, since the Goemans Williamson algorithm gives a solution which satisfies $(1 - \beta \sqrt{\delta})$ fraction of the constraints, given that $\text{OPT}(G) \geq 1 - \delta$, for an appropriate $\beta (\beta \geq 2\sqrt{\alpha GW} - \alpha GW$ or so). The UGC implies this conjecture, from [KKMO04].

The MAX-CUT instance $G$ is not necessarily bipartite (in fact, bipartite instances are easy), so one can not directly view it in the UG formulation (equivalently, as a 2-Prover 1-Round game). However, one can convert it into a bipartite UG instance $U$, as was shown in Section 2.1.1 with $\text{val}(U) = \frac{1}{2} + \frac{1}{2} \text{OPT}(G) = 1 - \frac{1}{2} \text{OPT}(G)$. Now, as the claim below shows, a strong enough parallel repetition would show that the MAX-CUT conjecture is equivalent to the UGC.

**Claim 4.15.** Let $\alpha$ be the minimum value such that for any 2-Prover 1-Round Unique game $G_U$, we have $\forall n \in \mathbb{N} : \text{val}(G_U^n) = (1 - \epsilon) \implies \text{val}(G_U^{\otimes n}) \leq (1 - \epsilon^\alpha)^{\Omega(n)}$. Then if $\alpha < 2$, then the MAXCUT conjecture implies the UGC.

**Proof.** We will apply parallel repetition to the UG instances corresponding to MAX-CUT; parallel repetition preserves uniqueness, so the resulting games are also UG instances. For the two classes of instances, we have:

1. (Completeness) $G \in \text{YES} \implies \text{OPT}(G) \geq 1 - \delta \implies \text{val}(G_U^n) \geq 1 - \frac{\epsilon}{2}$. On parallel repetition, we get $\text{val}(G_U^{\otimes n}) \geq (1 - \frac{\epsilon}{2})^n \geq 1 - \frac{n \delta}{2}$.

2. (Soundness) $G \in \text{NO} \implies \text{OPT}(G) \geq 1 - \beta \sqrt{\delta} \implies \text{val}(G_U) \leq 1 - \frac{\beta \sqrt{\delta}}{2}$. On parallel repetition, we get $\text{val}(G_U^{\otimes n}) \leq (1 - \frac{\beta \sqrt{\delta}}{2})^n \leq (1 - \frac{\beta \sqrt{\delta}}{2})^{\Omega(n)}$.

If $\alpha < 2$, then for $n \approx \delta^{-\alpha/2} \log(\frac{1}{\epsilon})$, we have that the second quantity is $(1 - \delta^{\alpha/2})^{\Omega(n)} \leq \delta$, and $(1 - \delta n/2) = 1 - O(\delta^{1 - \alpha/2} \log(1/\delta))$. Thus, as we shrink $\delta \rightarrow 0$, we will have that the completeness parameter $\rightarrow 1$ and the soundness parameter $\rightarrow 0$, and these instances will be NP-hard to distinguish between, since we started out with hard to distinguish MAX-CUT instances.

However, Raz's counterexample on the odd-cycle game actually says that one cannot get a better exponent than 2 for $\alpha$ above, blocking the amplification setup. In fact, as pointed out in Kho10, Theorem 4.12 itself suggests to us why the amplification would not have succeeded in the first place. For the YES cases, the completeness parameter is close to 1 for $n \delta \ll 1$. At the same time, the MAX-CUT conjecture...
tells us that an SDP-relaxation would not be able to distinguish between 
\(1 - \beta \sqrt{\delta}\) and \(1 - \delta\), hence there are NO instances where we would have the SDP value of the game as \(\sigma(G_U) \geq 1 - \delta\). But applying Theorem 4.12 gives us that \(\text{val}(G_{U}^{\otimes n}) \geq 1 - \sqrt{n\delta} \approx 1\). That is, there are NO instances with value very close to 1 under parallel repetition, rather than 0.

An alternative form of the UGC
Using the parallel repetition theorem for projection games, Rao [Rao08] proves that the following a-priori weaker statement is in fact equivalent to the UGC:

**Conjecture 4.16.** There is an increasing unbounded function \(\Gamma : \mathbb{R}^+ \rightarrow \mathbb{R}^+\) such that the following holds: for every \(\epsilon > 0\), there exists a constant \(k = k(\epsilon)\) such that given a UG instance \(U \in \text{UG}_{[k]}\), it is NP-hard to distinguish between the cases:

- **YES** = \(\{U : \text{OPT}(U) \geq 1 - \epsilon\}\)
- **NO** = \(\{U : \text{OPT}(U) \leq 1 - \sqrt{n\epsilon}\}\)

**Claim 4.17.** The above conjecture is equivalent to the UGC

**Proof.** That Conjecture 4.16 is implied by the UGC is obvious: setting \(\delta = \epsilon\) in the UGC would imply the hardness that Conjecture 4.16 asks for with \(\Gamma(1/\epsilon) = (1 - \epsilon)/\sqrt{\epsilon}\ (\epsilon < 1)\).

For the reverse implication, let us start with the given conjecture. Let \(n = \frac{1}{\Gamma^{2}(1/\epsilon)}\). Given an instance \(U\), we transform it using parallel repetition to another unique game, \(U^{\otimes n}\). This gives us:

- (Completeness) If \(U \in \text{YES}\), then \(\text{OPT}(U^{\otimes n}) \geq (1 - \epsilon)^n \geq 1 - n\epsilon = 1 - \frac{1}{\Gamma^{2}(1/\epsilon)}\)
- (Soundness) If \(U \in \text{NO}\), then \(\text{OPT}(U^{\otimes n}) \leq (1 - \Omega((\Gamma(1/\epsilon))^2))^{\frac{1}{\Gamma^{2}(1/\epsilon)}} \leq \exp(-\Omega(\Gamma(1/\epsilon)))\)

The completeness parameter tends to 1 and the soundness parameter tends to 0 as \(\epsilon \rightarrow 0\), implying the hardness that the UGC postulates.

We conclude this Section here, and move on to some related results that utilize the Parallel Repetition theorem.
5 Related Results

This section discusses some results related to the Parallel Repetition Theorem. We begin with a connection to a geometric problem, and then state some results for special classes of games.

5.1 Tiling in High Dimensions

In [FKO07], Feige et. al. show that the value of the $d$-repeated odd-cycle game we considered in Section 4.1 is closely related to the geometric problem of “tiling” the $d$-dimensional space $\mathbb{R}^d$ with bodies of unit volume, but minimal surface area. This problem is called as the foam problem in literature. We donot give a mathematically rigorous definition of the terms here, and hope that intuition will be enough to guide the main ideas.

**Definition 5.1.** A fundamental domain in $\mathbb{R}^d$ over $\mathbb{Z}^d$ is a compact body $K \subseteq \mathbb{R}^d$ with a piecewise “smooth” closed surface as boundary, and $\bigcup_{v \in \mathbb{Z}^d}(K + v) = \mathbb{R}^d$, with interior($K$) $\cap$ interior($K + v$) $= \emptyset$, $\forall v \in \mathbb{Z}^d$.

We say that the fundamental domain tiles $\mathbb{R}^d$ over $\mathbb{Z}^d$. The foam problem is to find a fundamental domain of minimum surface area. Note that a fundamental domain will always have unit volume. We will be looking at tilings that are periodic in $\mathbb{R}^d$ with respect to $\mathbb{Z}^d$. That, is, we are looking for a body with a closed, piecewise smooth surface, of unit-volume, that, under integer shifts, covers $\mathbb{R}^d$ with the interiors of each shifted version being disjoint. Note that in $\mathbb{R}^2$, the term volume translates as area, and surface area translates into circumference or perimeter. Some examples of tilings in $\mathbb{R}^2$ are shown in Figure 5.1.

![Figure 5.1: Tiling $\mathbb{R}^2$ using “bricks”. The associated spine is shown on the left.](image)

It turns out that foams are closely related to objects on the unit-torus called spines. A torus in $d$-dimensions is defined as $T^d = \mathbb{R}^d/\mathbb{Z}^d$; in $\mathbb{R}^2$, it is the unit square, with opposite sides identified (see Fig. 5.2). A spine on a torus is defined as a $(d - 1)$-dimensional surface that intersects every nontrivial cycle on the torus. We will only be considering piecewise smooth spines here. Throughout, we will let $A(d)$ stand of the minimum surface area of a smooth spine in $T^d$.

The following questions turn out be essentially the same: What is the least surface area of the boundary of a fundamental domain in $\mathbb{R}^d$, and what is the least surface area of a spine in $T^d$.

**Fact 5.2.** The minimum surface area of a fundamental domain in $\mathbb{R}^d$ is $2 \times A(d)$.

---

2 A closed surface is one that is compact, and without a boundary
3 A set of points in the interior of $K$, but not contained in $K$
4 A nontrivial cycle is a closed loop that cannot be shrunk to a point
In $\mathbb{R}^2$, we can visualize why this should be true by taking the torus (i.e. the unit square) with the spine, and tiling it across the plane. A non-trivial loop in the torus now emerges as a infinite path in $\mathbb{R}^2$, and the whole space is not covered by the induced tiling unless some part of the spine “blocks” this infinite path. (See Fig. 5.3).

For any $d$, the isoperimetric inequality implies that since the sphere has the minimal surface area for any given volume, we will have that $A(d) \geq O(\sqrt{d})$.

**Definition 5.3.** (Cycle class) The class of a cycle (i.e. closed loop) $\gamma$ on $T^d$ is a $d$-dimensional vector, denoted by $[\gamma] = (\gamma_1, \ldots, \gamma_k)$, where $\gamma_i$ denotes the number of times the loop goes around co-ordinate axis $i$. Note that a cycle is nontrivial iff $[\gamma] \neq 0$. A cycle is called topologically odd, if $\gamma_i$ is odd for some $i$.

### 5.1.1 From Parallel Repetition to Cycle Elimination

First, we need to define the discrete torus graph

**Definition 5.4.** (Discrete torus graph) For $d, m \in \mathbb{N}$, the discrete torus graph with $l_\infty$-edge structure, denoted by $(Z^d_m)_\infty$ is the graph on the vertex set $Z^d_m$, where two vertices are connected iff their $l_\infty$ distance is atmost 1. Here, $Z^d_m$ is simply every co-ordinate of $z \in Z^d$ reduced modulo $m$.

A slightly modified graph will be more convenient for us:
Proof. The idea is to consider the vertices in the graph $K_2 \times (\mathbb{Z}_m^d)_{\infty}$ as questions to the provers, where $P1$ gets $(0, x)$ and $P2$ gets $(1, y)$. Due to the fact that $(0, x), (1, x)$ gets asked to the provers with probability $\frac{1}{2}$ of all other edges of the form $\{(0, x), (1, y)\}$ for $x \neq y$, we will count such edges with weight $\frac{1}{2}$.

Observe that the edges in the graph corresponds to questions that can get asked to the provers. Now, for any strategy $S$ of the provers, suppose they err on a set of edges $E(S)$. Obviously, the probability of winning with strategy $S$ is $1 - \frac{|E(S)|}{\frac{3}{2}m^d}$, where the total number of edges is $3^d m^d$. What follows will show that every strategy $S$ of the two provers must be odd-cycle blocking, and vice versa.

(1. Strategy $\implies$ Blocking) Suppose strategy $S$ is not blocking. This means that there is a co-ordinate $i$ around which a cycle $\gamma_S = ((0, x_1), (1, y_1), \ldots, (1, y_n), (0, x_1))$ goes around an odd number of times. Look at the projection of the vertices involved in this cycle onto co-ordinate $i$, giving us $((0, x_1[i]), (1, y_1[i]), \ldots, (0, x_1[i]))$. The corresponding cycle in $(\mathbb{Z}_m^d)_{\infty}$ when projected onto the $i$-th co-ordinate looks like $(x_1[i], y_1[i], \ldots, x_1[i])$. Since these are questions answered correctly, the colorings...
assigned to the corresponding vertices in $K_2 \times (\mathbb{Z}_m^d)_{\infty}$ in co-ordinate $i$, flips at every 1-step in this (i.e. $x_j[i] \neq y_j[i]$ or $y_j[i] \neq x_{j+1}[i]$). However, since $m$ is odd, and $|\gamma|_i$ is odd, it means that the number of 1-steps is odd. This would mean that the colour assignment to the first and the last vertex are different, but this is impossible since being a cycle, both are from the same vertex $(0, x_1)$.

(Part 2: Blocking $\Rightarrow$ Strategy) Consider a set of edges $\mathcal{E}$ that are blocking. Look at a connected component $C$ of $K_2 \times (\mathbb{Z}_m^d)_{\infty} \setminus \mathcal{E}$. Fixing a colouring of any vertex $v$ in $C$ will then uniquely fix the colors of all other vertices in that component (the game is a unique game). This procedure will get stuck only if there is a cycle that cannot be labelled consistently. However, this happens only when the cycle is topologically odd in some co-ordinate $i$, but since $E$ is blocking, this cannot happen.

Now we translate back cycles in $K_2 \times (\mathbb{Z}_m^d)_{\infty}$ to cycles in $(\mathbb{Z}_m^d)_{\infty}$:

**Theorem 5.9.** Let $\delta(d, m)$ and $\delta'(d, m)$ be defined as in Problems 5.7 and 5.6. Then $\delta'(d, m) \leq 2\delta(d, m)$

**Proof.** Let $\mathcal{E}$ in $(\mathbb{Z}_m^d)_{\infty}$ block all non-trivial cycles. We will construct an edge set in $K_2 \times (\mathbb{Z}_m^d)_{\infty}$ that will block all topologically non-trivial cycles. For every $(x, y)$ in $\mathcal{E}$, add $\{(0, x), (1, y)\}$ and $\{(0, y), (1, x)\}$ to $\mathcal{E}'$. Then $|\mathcal{E}'| = 2|\mathcal{E}|$, and $\mathcal{E}'$ blocks all nontrivial cycles on $K_2 \times (\mathbb{Z}_m^d)_{\infty}$, in particular topologically odd cycles.

Thus, combining the previous two theorems, we get:

$$\text{val}(\mathcal{G}^d_{C_m}) \geq 1 - 2\delta(d, m) \quad (5.1)$$

We are almost at relating the value of the odd cycle game to the spine (and hence, foam) problem. Note that a spine on $\mathbb{T}^d$ blocks all nontrivial cycles (including continuous ones). The graph $(\mathbb{Z}_m^d)_{\infty}$ when embedded in the torus, has a blocking strategy that blocks only some discrete non-trivial cycles (those that result from edges of the graph itself). The next theorem relates these two quantities.

**Theorem 5.10.** Let $A$ be the surface area of any spine in $\mathbb{T}^d$, then $\delta(d, m) \leq \sqrt{2/3}A/m$. Thus, if $A(d)$ is the minimum area of a spine, then $\delta(d, m) \leq \sqrt{2/3}A/d/m$.

We donot give the proof of this theorem here, and refer the reader to [FKO07] for the proof. The proof is probabilistic, and the basic idea is as follows: we start with a spine $\mathcal{S}$, and randomly translate it in $\mathbb{T}^d$. If we delete every edge that $\mathcal{S}$ intersects in $(\mathbb{Z}_m^d)_{\infty}$, then we block all non-trivial cycles in $(\mathbb{Z}_m^d)_{\infty}$. We then calculate the expected number of edges a random translate intersects, and show that this is atmost the RHS.

Given the above relations, we see that one could prove a lower bound on the value of the repeated odd-cycle game, by finding a spine (equivalently, a tiling) of minimal surface area.

### 5.1.2 Spine construction

![Figure 5.5: The translation procedure to produce the spine required, in [KORW08]. (a) A level set of $f$ (b) 3 random translations (c) Partial spine (shown in red) output by the 3 translations](image-url)

Figure 5.5: The translation procedure to produce the spine required, in [KORW08]. (a) A level set of $f$ (b) 3 random translations (c) Partial spine (shown in red) output by the 3 translations
Kindler et. al. [KORW08] show that a spine exists with surface area atmost $2\pi \sqrt{d}$, so we have that $A(d) \leq 2\pi \sqrt{d}$. Note that this body is of unit volume, but it’s surface area is asymptotically like that of a sphere. We do not give the proof of their result here, but describe the rough outline of the steps used.

We can represent a spine $S$ (in general, any $S \subset \mathbb{T}^d$) by a $0-1$ valued function $f : \mathbb{T}^d \to \{0,1\}$ such that $f(x) = 0$ for $x \in S$ and 1 otherwise. The surface area of such a spine is then $\int \|\nabla f\|$.

What we are looking for is then a function $f : \mathbb{T}^d \to \{0,1\}$ satisfying the following:

1. $\int f = 1$
2. $\{x : f(x) = 0\}$ is a spine
3. $\int \|\nabla f\|$ is minimal

It turns out that we can relax the discreteness criterion on $f$, satisfying the above conditions, and still manage to get a small spine (note that when discreteness is relaxed, point 3 need not necessarily imply that the spine given by the zeroes of $f$ is small). Given a continuous $f$, the spine is constructed by considering random translates by $Z$ of $f$ in $\mathbb{T}^d$ and unioning the boundaries of the level sets $f(x - Z) \geq T$ ($T$ is randomly chosen and fixed before the translations are applied). It can be proven that this randomized construction results in a spine, with expected surface area $\int \|\nabla f\|$.

The next step is to relax criterion 2 above to restrict $f$ to represent the special class of spines which vanish at the boundary of the region; i.e. $f(x_1 \ldots x_d) = 0$ whenever any $x_i = 0$. This, the authors show, is enough to get a small spine. So, all we want is a continuous function of unit volume in $\mathbb{T}^d$ that vanishes on the boundary points, and has $\int \|\nabla f\|$ as small as possible. In order to do this, the function that is given is: $f = g^2$ where

$$g(x) = \prod_{i=1}^{d} \sqrt{2 \sin(\pi x_i)}$$

For this particular choice of $g$, we will have $\int \|\nabla f\| \leq 2\pi \sqrt{d}$. This shows that $A(d) \leq 2\pi \sqrt{d}$, and hence $\text{val}(G^{\otimes d}_{C_{\text{ma}}}) \geq 1 - 2\pi \sqrt{2/3} \frac{\sqrt{d}}{\pi}$, which matches Raz’s strategy in Section 4.1.

Subsequently, Alon and Klartag [AK08] gave an alternative, simpler proof for the above problem, that achieves the same values for $A(d)$. For the discrete case, they use Cheeger’s inequality and a random translation scheme similar to the one given above [KORW08] to conclude the existence of small discrete spines, without going into the continuous domain.

5.2 Other Results

In this concluding section, we very briefly survey some results related to the technique of parallel repetition. We leave out the proofs of these results for lack of space, and only give very brief ideas about the motivation and implications for each of these.

5.2.1 Strong Parallel Repetition for Free Projection games

In [BRR+09], Barak et. al. show that strong parallel repetition does hold for a subclass of projection games that are also free.

**Definition 5.11.** (Free Games) A 2-Prover 1-Round game is called a free game, if the questions to the two players are a product distribution. That is, $\Pr_{D}[X = x \land Y = y] = \Pr_{X}[X = x] \Pr_{Y}[Y = y]$.

The results in the paper are the following:

**Theorem 5.12.** For a free game $G$ with $\text{val}(G) = 1 - \epsilon$, we will have $\text{val}(G^{\otimes n}) \leq (1 - \epsilon/9)^{n/(18c+3)}$.

**Theorem 5.13.** For a free projection game $G$ with $\text{val}(G) = 1 - \epsilon$, $\text{val}(G^{\otimes n}) \leq (1 - \epsilon/9)^{(n/33) - 1}$

The $f$ used here is not differentiable, so one can use the limit of a sequence of differentiable functions $f_i$ that tend to $f$. 

41
Following the line of reasoning of the proof for the general case presented in Section 5.2.3, assume that for some $S \subseteq [n]$, $\Pr[W_S]$ is high, then conditioned on $W_S$, on an average co-ordinate outside $S$, the distribution of questions is similar to the questions of a single round game.

However, the main observation in is the fact that for the special case of free games, this difference can be viewed in the relative entropy world, rather than using the $L_1$ norm. Further, since there is no dependence between $X_i, Y_i$, except due to the conditioning on $W_S$, the random variables drawn in correlated sampling ($A^{-i}$ in the general case) will simply be the questions and answers in $S$: $X^k, Y^k, A^k$ (conditioned on $W_S$) (there are some subtleties here, which we ignore for the current discussion). Thus, the main lemma proved is of the following form:

**Lemma 5.14.** (Main Lemma for free games) Let $S = \{n-k+1, \ldots, n\}$. Then, for a free game $G$, we have for the game $G^{\otimes n}$, the following holds:

$$
\mathbb{E}_{i \in [n-k]} \mathbb{E}_{X^k Y^k A^k | W_S} D((X_i Y_i | (X^k Y^k A^k) | W_S) \parallel \{X_i Y_i\}) \leq \frac{1}{n-k} (k \log |A| - \log \Pr[W_S])
$$

(5.2)

We haven’t yet saved on much, since $D(P \parallel Q) \geq \|P - Q\|_1^2$, and we had used the fact that $\|P - Q\|_1 = \max_{E \text{ event}} |P(E) - Q(E)|$. However, the following fact can be used:

**Fact 5.15.** For all probability distributions $P, Q$ over the same space $\Omega$, and for every event $T \subseteq \Omega$, if $D(P \parallel Q) \leq \delta$ and $P(T) \leq \delta$ then $Q(T) \leq 4\delta$.

By an appropriate setting of parameters: $\Pr[W_S] \geq 2^{-\epsilon(n-k)/9 + kc}$, we have that the RHS of (5.2) is $\leq \epsilon/9$. Now, suppose that for all $i$, $\Pr[W_i | W_S] > 1 - \epsilon/9$. This means $\Pr[W_i | W_S] < \epsilon/9$. This can be rewritten as $\mathbb{E}_{i \in [n-k]} \mathbb{E}_{X^k Y^k A^k | W_S} [\Pr[W_i | (X^k, Y^k, A^k) \wedge W_S] < \epsilon/9$, for all $i \in [n-k]$ . This gives us that

$$
\mathbb{E}_{i \in [n-k]} \mathbb{E}_{X^k Y^k A^k | W_S} \left( D((X_i Y_i | (X^k Y^k A^k) \wedge W_S) \parallel \{X_i Y_i\}) + \Pr[W_i | (X^k Y^k A^k \wedge W_S)] \right) < 2\epsilon/9 < \epsilon/4
$$

This means that there is a co-ordinate $i$ and $(X^k = x^k, Y^k = y^k, A^k = a^k)$ such that:

$$
\Pr[W_i | (X^k Y^k A^k) = (x^k y^k a^k) \wedge W_S] < \epsilon/4
$$

$$
D((X_i Y_i | (X^k Y^k A^k) = (x^k y^k a^k) \wedge W_S) \parallel \{X_i Y_i\}) < \epsilon/4
$$

Now we invoke fact 5.15 to conclude that $\Pr[W_i] < 4 \times \epsilon/4 = \epsilon$, which is a contradiction, since $\Pr[W_i] = 1 - \epsilon$. Thus, if $\Pr[W_S] \geq 2^{-\epsilon(n-k)/9 + kc}$, then there is an $i \in [n-k]$ such that $\Pr[W_i | W_S] \leq 1 - \epsilon/9$. The rest of the proof goes through by induction as in the general case.

For the case of projection games, using arguments similar to Rao’s argument \cite{Rao08}, the dependence on answer set size is eliminated, and we save a factor of $\epsilon$. We omit the details.

### 5.2.2 2-query PCPs

We had seen in Section 5.2.2 how parallel repetition is a useful tool to reduce the soundness parameter in 2-query PCPs. A recent result by Impagliazzo et al. \cite{IKW09} gives a slightly different method to reduce soundness error in 2-query PCPs.

Given a 2-query PCP over the alphabet $\Sigma$, i.e. $PCP_{c,s}[\tau(n), 2]$ for a language $L \subseteq \{0, 1\}^*$, one can view it as a constraint graph for every $x \in \{0, 1\}^*$. The vertices of the graph $G_x$ correspond to locations in the proof supplied for $x$, and there is an edge between 2 vertices $(u, v)$, if they can be queried by the verifier simultaneously while checking. Every edge has a constraint $\phi_e$ that describes when the verifier can accept on that view. Thus, we will, in this section, speak about constraint graphs $(G, \Phi)$ interchangeably with 2-query PCPs.

Given such a PCP for some $L$, consider the following proof checking system (again for $L$):

**Verifier $\mathcal{Y}$:**

1. Let $k' = \Theta(k)$. Pick the following random sets:

42
A set of random vertices of size \( k' \), call it \( A \).

For each \( v \in A \), pick a random edge \((v, v') \in G \) and call \( A_{E1} \) this set of edges.

Independently, repeat step 1(b), and call these edges \( A_{E2} \).

Pick two random sets of edges of size \( k - k' \) each, call them \( B_{E1}, B_{E2} \)

2. Query \( C_{E}(A_{E1}, B_{E1}) \), \( C_{E}(A_{E2}, B_{E2}) \) and accept iff:
   
   (a) Atleast 0.9c fraction of constraints are passed on \( B_{E1}, B_{E2} \), and
   
   (b) The assignments are consistent on the set \( A \)

Here, the proof the verifier expects is assignments to \( k \)-sets of edges in the graph. The following can then be shown:

**Theorem 5.16.** If \( x \in L \), (i.e. the constraint graph for \( x \) is \( c \)-satisfiable) then there is a proof \( C_{E} \) accepted by \( \mathcal{Y} \) with probability \( c' \geq 1 - \exp(-ck) \). Further, if \( c = 1 \), then \( c' = 1 \).

If \( x \notin L \), (i.e. constraint graph for \( x \) is \( s \)-unsatisfiable) then no proof \( C_{E} \) is accepted by the verifier with probability greater than \( s' = \exp(-c\delta k') = \exp(-c\delta \sqrt{k}) \) provided \( s' \leq 1/4 \), where \( c > 0 \) is a constant.

If the original proof is of size \( n \), then the new proof is of size \( \approx (n^2)^k = O(n^k) \), and the alphabet is \(|\Sigma|^{2k}\).

By using methods similar to those required in the proof of the above theorem, the authors give an improved analysis of the parallel repetition theorem of Feige and Kilian for the case of miss-match games that we saw in Section 2.4. For the two prover case, there are two logical parts to the proof, call the proof for player 1 \( C_{1} \) and that for player 2 \( C_{E} \). The verifier does the following (given a constraint graph \((G, \Phi)\)):

**Verifier \( \mathcal{Y}' \):**

1. Let \( k' = \Theta(k) \). Pick a set of \( k' \) random vertices \( A \), and for each vertex \( v \in A \), pick a random incident edge in \( G \). Call the set of these edges \( A_{E2} \). (These are the match rounds)

2. Pick a set of \( k - k' \) random edges \( B_{E1} \). (Parts of these will correspond to the miss rounds)

3. Query \( C_{1}(A) \) and \( C_{E}(A_{E2}, B_{E2}) \). Accept iff:
   
   (a) the query satisfies \( \geq 0.9c \) fraction of constraints of \( B_{E2} \),
   
   (b) the assignment to \( A \) is consistent.

Here, \( C_{1} \) is expected to contain the labellings for all subsets of vertices of size \( k \). \( C_{E} \) is expected to contain labellings for all size-\( k \) subsets of edges. We note that verifier \( \mathcal{Y}' \) has the projection property. The following can be said about \( \mathcal{Y}' \).

**Theorem 5.17.** (Completeness) If the constraint graph is \( c \)-satisfiable, then there is a pair of proofs \((C_{1}, C_{E})\) such that \( \mathcal{Y}' \) accepts with probability \( c' \geq 1 - \exp(-ck) \).

(Soundness) If the constraint graph is less than \( s \)-satisfiable, then no proof \((C_{1}, C_{E})\) is accepted with probability greater than \( s' = \exp(-c\delta \sqrt{k}) \), provided \( s' \leq 1/4 \).

The authors leave it open, if the rate of decay in soundness can be made exponential in \( k \), rather than \( \sqrt{k} \).

In this survey, we have not covered all results related to the applications of 2-Prover 1-Round games, we end by indicating two areas that we have skipped. There are results where the underlying constraint graph is an expander [SS07], [AKK+08], [KR10], which allows stronger parallel repetition theorems. Also, 2-Prover 1-Round games with entanglement behave quite differently than their classical counterparts: for instance, one can easily find the approximate value of unique games with entangled provers [KRT08, KR10].
References


A Hellinger Distances

Here we give proofs of the properties of Hellinger Distances we used, for the case when the distributions are over a finite domain. For distributions $D_1$ on $\Omega_1$ and $D_2$ on $\Omega_2$, define $D_1 \otimes D_2$ to be their product distribution over the domain $\Omega_1 \times \Omega_2$.

Definition A.1. (Hellinger distance) The Hellinger distance between two probability distributions $P, Q$ over a domain $\Omega$ is defined by

$$H^2(P, Q) = \frac{1}{2} \sum_{\omega \in \Omega} (\sqrt{P(\omega)} - \sqrt{Q(\omega)})^2$$

- Let $|\Omega| = d$, so that we identify $\Omega$ with $[d]$. Then any distribution $P$ on $\Omega$ can be viewed as a vector in $\mathbb{R}^d$. Let $\vec{p} \in \mathbb{R}^d$ be a vector given by $p_i = \sqrt{P(i)}$. We note that $\|\vec{p}\| = 1$. In general, let $\phi$ be a map from distributions $D$ to their corresponding unit-norm vectors $\vec{d}$. Then $H^2(P, Q) = \frac{1}{2}\|\phi(P) - \phi(Q)\|^2$.

Claim A.2. For $P, Q$ being some distributions over $\Omega$, $H^2(P, Q) \leq \|P - Q\|_1 \leq \sqrt{2}H(P, Q)$

Proof. Let $\phi(P) = u$ and $\phi(Q) = v$. Then we have:

$$\|P - Q\|_1^2 = \frac{1}{4} \left( \sum_{i \in [d]} |u_i^2 - v_i^2| \right)^2$$

$$= \frac{1}{4} \left( \sum_{i \in [d]} |u_i + v_i||u_i - v_i| \right)^2$$

$$\leq \frac{1}{4} \left( \sum_{i \in [d]} |u_i + v_i|^2 \right) \left( \sum_{i \in [d]} |u_i - v_i|^2 \right) \quad \text{(using Cauchy Schwarz)}$$

$$\leq \frac{1}{4} (\|u\|^2 + \|v\|^2) \times 2H^2(P, Q)$$

$$= 2H^2(P, Q)$$

For the inequality on the LHS, note that all entries in $u, v$ are non-negative. For each $i$, let $M_i = \max(u_i, v_i)$ and $m_i = \min(u_i, v_i)$. Then we have:

$$\|P - Q\|_1 = \frac{1}{2} \sum_{i \in [d]} (M_i^2 - m_i^2)$$

$$\geq \frac{1}{2} \sum_{i \in [d]} (M_i - m_i)^2 \quad \text{(since } M_i + m_i \geq M_i - m_i\text{)}$$

$$= H^2(P, Q)$$

Claim A.3. Let $P_1, Q_1$ be distributions over $\Omega_1$, and let $P_2, Q_2$ be distributions over $\Omega_2$. Then, we have

$$H^2(P_1 \otimes P_2, Q_1 \otimes Q_2) \leq H^2(P_1, Q_1) + H^2(P_2, Q_2)$$

Proof. Let $u_i = \phi(P_i)$, $v_i = \phi(Q_i)$, for $i \in \{1, 2\}$. We note that $\phi(P_1 \otimes P_2) = u_1 \otimes u_2$ (the Kronecker product). The given statement is proven, if we prove that

$$\|u_1 \otimes u_2 - v_1 \otimes v_2\|^2 \leq \|u_1 - v_1\|^2 + \|u_2 - v_2\|^2$$

46
Consider the two orthogonal subspaces: $S_1$ spanned by $\frac{u_1 + v_1}{2} \otimes R^{[\Omega_2]}$, and its orthogonal complement $S_2$. Projecting the vector on the LHS (call it $w$) onto $S_1$ gives (using $\|u\| = \|v\|$):

$$w_1 = \frac{u_1 + v_1}{2} \otimes (u_2 - v_2)$$

Similarly, projecting it onto $S_2$ gives the vector :

$$w_2 = \left( u_1 - \frac{(u_1 + v_1)}{2} \right) \otimes u_2 - \left( v_1 - \frac{(u_1 + v_1)}{2} \right) \otimes v_2$$

$$= \frac{(u_1 - v_1)}{2} \otimes u_2 + \frac{(u_1 - v_1)}{2} \otimes v_2$$

$$= \frac{(u_1 - v_1) \otimes (u_2 + v_2)}{2}$$

Noting that $\|w\|^2 = \|w_1\|^2 + \|w_2\|^2$, and the fact that $\|\frac{u_1 + v_1}{2}\|^2 \leq 1$, the result follows.
B  Multiplicativity of SDP relaxations for games

Here we give a proof of multiplicativity of the SDP formulation for 2-Prover 1-Round games, considered in Section 4.2.3. The discussion here follows that in [KR10], which is very similar to the treatment in [FL92]. Throughout this discussion, and in fact in the main body of the report too, we will represent SDP’s using vector programs; that is, an optimization problem with vectors as variables, and all the objectives and constraints are linear combinations of the inner products of pairs of vector variables. As is well known, the standard SDP form and this form are equivalent.

Definition B.1. (Bipartite SDPs) A bipartite SDP involves 2 distinct sets of vector variables: \( U = \{u_1 \ldots u_n\} \) and \( V = \{v_1 \ldots v_n\} \). The objective function of the SDP involves only inner products of the form \( \langle u_i, v_j \rangle \), where \( u_i \in U \) and \( v_j \in V \). Further, every constraint is an equality constraint and is expressed as a linear combination of inner products in either \( U \) or \( V \) completely. That is, every constraint is of the form: \( \sum_{ij} A_{ij} \langle w_i, v_j \rangle = a \), where either \( \forall i : w_i \in U \) or \( \forall i : w_i \in V \). More specifically, it has: a \( n_1 \times n_2 \) symmetric matrix \( J \), \( L_1 \) symmetric matrices \( A^1 \ldots A^{L_1} \) of dimension \( n_1 \times n_1 \), \( L_2 \) symmetric \( n_2 \times n_2 \) matrices \( B^1 \ldots B^{L_2} \), and real numbers \( a_1 \ldots a_{L_1}, b_1 \ldots b_{L_2} \). The SDP program (denoted by \( S \)) is:

\[
\begin{align*}
\text{max} & \quad \sum_{i \in [n_1], j \in [n_2]} J_{ij} \langle u_i, v_j \rangle \\
\text{subject to} & \quad \sum_{i, j \in [n_1]} A^l_{ij} \langle u_i, u_j \rangle = a_l \quad \forall l \in \{1 \ldots L_1\} \\
& \quad \sum_{i, j \in [n_2]} B^l_{ij} \langle v_i, v_j \rangle = b_l \quad \forall l \in \{1 \ldots L_2\}
\end{align*}
\]

It is easy to check that the SDP formulation for 2-Prover 1-Round games is a bipartite SDP. To look at the repeated game, we define the bipartite product of two bipartite SDPs. All the SDPs considered henceforth in this section will be bipartite, though we will not explicitly mention it.

Definition B.2. (Bipartite product) Given two SDPs \( S(U, V, J, \{A^l\}_{l \in [L_1]}; \{B^l\}_{l \in [L_2]}; \{a^l\}_{l \in [L_1]}; \{b^l\}_{l \in [L_2]} \) and \( S'(U', V', J', \{A'^{l'}\}_{l' \in [L_1']}; \{B'^{l'}\}_{l' \in [L_2']}; \{a'^{l'}\}_{l' \in [L_1']}; \{b'^{l'}\}_{l' \in [L_2']} \) the bipartite product \( S \otimes S' \) is specified by the bipartite SDP \( R \) with \( n_1 n_1' + n_2 n_2' \) variables, \( L_1 L_1' + L_2 L_2' \) constraints (the constraint matrices are \( \{A^l \otimes A'^{l'}\} \) and \( \{B^l \otimes B'^{l'}\} \)), \( n_1 n_1' \times n_2 n_2' \) objective matrix \( J \otimes J' \) and real numbers \( a_l a_{l'} \) and \( b_l b_{l'} \).

\[
\begin{align*}
\text{max} & \quad \sum_{i=1, j=1', k=1, k'=1}^{n_1, n_1'} J_{hh'} \langle u_{ih}, v_{ij} \rangle \\
\text{subject to} & \quad \sum_{i=1, j=1, k=1}^{n_1, n_1'} A^l_{ij} A'^{l'}_{h'k} \langle u_{ih}, u_{jk} \rangle = a_l a_{l'} \quad \forall l \in \{1 \ldots L_1\}, \forall l' \in \{1 \ldots L_1'\} \\
& \quad \sum_{i=1, j=1, k=1}^{n_2, n_2'} B^l_{ij} B'^{l'}_{h'k} \langle v_{ih}, v_{jk} \rangle = b_l b_{l'} \quad \forall l \in \{1 \ldots L_2\}, \forall l' \in \{1 \ldots L_2'\}
\end{align*}
\]

Observe that if \( G \) has SDP relaxation \( S \), then \( G^{\otimes 2} \) has SDP relaxation \( S \otimes S \). For any SDP program \( S \), denote by \( \omega(S) \) it’s optimal value. We will need the following easy-to-prove fact:

Fact B.3. Given two symmetric matrices \( X, Y \), if \( X \geq Y \) then for any other matrix \( R \), we have \( R^T X R \geq R^T Y R \).

We will also use the following crucial fact regarding duality of bipartite SDPs, which we won’t provide a proof of here. Call a SDP strictly feasible, if there is a feasible solution where all the vectors are non-zero.

Fact B.4. If a bipartite SDP \( S \) is strictly feasible, then strong duality holds, i.e. the optimal value of the dual is equal to \( \omega(S) \).
One can verify that in the case of the SDP for 2-Prover 1-Round games, strict feasibility is satisfied. Also, we note that for the 2-Prover 1-Round game, the optimal value is strictly positive for non-trivial games.

**Theorem B.5.** For two strictly feasible SDPs $S$ and $S'$, if $\omega(S) > 0$ and $\omega(S') > 0$, then $\omega(S \otimes S') = \omega(S) \times \omega(S')$

**Proof.** It is easy to see that $\omega(S \otimes S') \geq \omega(S) \times \omega(S')$. Given solutions $\{u_i\}_{i=1}^{n_1}, \{v_j\}_{j=1}^{n_2}$ and $\{u'_i\}_{i=1}^{n'_1}, \{v'_j\}_{j=1}^{n'_2}$ to $S$ and $S'$, we see that the set $\{u_i \otimes u'_i\}$ and $\{v_j \otimes v'_j\}$ form feasible solutions to the bipartite product $S \otimes S'$. Further, the value of the SDP objective at this point is $\omega(S) \times \omega(S')$.

To prove the other side of the inequality, we look at the duals of these SDPs. The dual of $S$ (call it $D$) is given by:

$$
\min \sum_{i=1}^{L_1} x_i a_i + \sum_{i=1}^{L_2} y_i b_i \tag{B.1}
$$

subject to

$$
\left( \begin{array}{cc}
\sum_{i=1}^{L_1} a_i^T & 0 \\
0 & \sum_{i=1}^{L_2} y_i b_i^T
\end{array} \right) \succeq \left( \begin{array}{cc}
0 & J/2 \\
J/2 & 0
\end{array} \right) \tag{B.2}
$$

Here, the variables are simply scalars $\{x_i\}, \{y_i\}$ and the 0’s in the matrices represent the all-zeroes matrix of appropriate dimensions. The dual of $S \otimes S'$ (call it $D'$) is:

$$
\min \sum_{i=1}^{L_1,L'_1} x_i a_i a_i'^T + \sum_{i=1}^{L_2,L'_2} y_i b_i b_i'^T \tag{B.3}
$$

subject to

$$
\left( \begin{array}{cc}
\sum_{i=1}^{L_1,L'_1} a_i^T \otimes A'^T & 0 \\
0 & \sum_{i=1}^{L_2,L'_2} y_i b_i b_i'^T \otimes B'^T
\end{array} \right) \succeq \left( \begin{array}{cc}
J \otimes J'/2 & 0 \\
0 & J \otimes J'/2
\end{array} \right) \tag{B.4}
$$

Similarly, call the dual of $S'$ as $D'$. Then, using Fact B.4 the optimal value of $S$ is equal to the optimal value of $D$, call this value $\omega_S$. Similarly, call the optimal value of $S'$ and $D'$ as $\omega_{S'}$. Then, the following claim holds:

**Claim.** If the optimal solution for $S$: $\{x_i\}_{i=1}^{L_1}, \{y_i\}_{i=1}^{L_2}$ has $\omega_S > 0$, we will have some feasible (optimal) solution $\{\tilde{x}_i\}, \{\tilde{y}_i\}$ satisfying $\sum_i x_i a_i = \sum_i y_i b_i = \frac{\omega_S}{2}$. Similarly for $S'$.

**Proof.** Given the optimal solution, we consider the new variables $\tilde{x}_i = \alpha x_i$ and $\tilde{y}_i = \frac{\alpha}{\sqrt{\alpha}} y_i$, where $\alpha > 0$. The new scalars are again a feasible solution, since we can set $R = \text{diag}(\sqrt{\alpha}, \ldots, \sqrt{\alpha}, 1/\sqrt{\alpha}, \ldots, 1/\sqrt{\alpha})$ and use Fact B.3 to conclude feasibility of the constraint (B.2).

Denote $\sum_i x_i a_i = s_x$ and $\sum_i y_i b_i = s_y$. First, since $\omega_S > 0$, we cannot have that $s_x < 0$. This is because, setting $\alpha = \sqrt{|s_y|/|s_x|}$ gives a feasible solution with objective value $-\sqrt{|s_x||s_y|} + \sqrt{|s_y|s_x|} = 0$, which is less than $\omega_S$. Similarly, $s_y > 0$.

Now, setting $\alpha = \sqrt{s_y/s_x}$ returns an objective value of $2\sqrt{s_x s_y}$, which is atmost (and hence, equal to) $\omega_S$ (AM-GM). Thus, the constructed $\tilde{x}_i = \alpha x_i$, and $\tilde{y}_i = \frac{\alpha}{\sqrt{\alpha}} y_i$ are optimal, and satisfy the property in the given claim. \hfill \Box

Using the above claim, let us propose points for the dual of $S \otimes S'$. Suppose $\{x_i\}, \{y_i\}$ are feasible optimal solutions to $S$, and $\{x'_i\}, \{y'_i\}$ are feasible optimal solutions to $S'$ such that:

$$
\sum_{i=1}^{L_1} x_i a_i = \sum_{i=1}^{L_2} y_i b_i = \frac{\omega_S}{2} \tag{B.5}
$$

$$
\sum_{i=1}^{L'_1} x_i a_i' = \sum_{i=1}^{L'_2} y_i b_i' = \frac{\omega_{S'}}{2} \tag{B.6}
$$
Then, set \( \{x_{l'}\}_{l'=1}^{L_1} \) and \( \{y_{l'}\}_{l'=1}^{L_1} \). For this setting, the objective function of \( Q \) attains the value:

\[
v = 2 \left( \sum_{l} x_{l} a_{l} (\sum_{l'} x_{l'} a_{l'}) + (\sum_{l} y_{l} b_{l} (\sum_{l'} y_{l'} b_{l'}) \right)
\]

\[
= 2 \left( \frac{\omega_{S} \omega_{S'}}{4} + \frac{\omega_{S} \omega_{S'}}{4} \right) \quad \text{(using (B.5), (B.6))}
\]

All that remains for us is to show that the points \( \{x_{l'}\}, \{y_{l'}\} \) are feasible in \( Q \). To do this, we will need the following fact:

**Fact B.6.** If \( X,Y,X',Y' \) are all positive semidefinite, and further \( X \succeq Y, X' \succeq Y', X' \preceq -Y' \) then \( X \otimes X' \succeq Y \otimes Y' \).

**Proof.** The proof follows by observing that

\[
2(X \otimes X' - Y \otimes Y') = (X - Y) \otimes (X' + Y') + (X + Y) \otimes (X' - Y'),
\]

and the RHS is p.s.d. since it is a sum of p.s.d. matrices.

We also observe that the constraint (B.2) for \( S \) continues to hold if we multiply the right hand side matrix \(-1\). This corresponds to using Fact (B.3) with \( R = \text{diag}(1 \ldots 1, -1 \ldots -1) \). We can now use Fact B.6 on the matrices in constraint (B.2) for \( S \) and the corresponding constraint in \( S' \) to obtain:

\[
\begin{pmatrix}
(\sum_{l=1}^{L_1} x_{l} A_{l}^i) \otimes (\sum_{l=1}^{L_1} x_{l'} A_{l'}^i) & 0 & 0 & 0 \\
0 & \ast & 0 & 0 \\
0 & 0 & \ast & 0 \\
0 & 0 & 0 & (\sum_{l=1}^{L_2} y_{l} B_{l}^i) \otimes (\sum_{l=1}^{L_2} y_{l'} B_{l'}^i)
\end{pmatrix}
\succeq
\begin{pmatrix}
0 & 0 & J \otimes J'/4 & 0 \\
0 & 0 & 0 & 0 \\
0 & J \otimes J'/4 & 0 & 0 \\
J \otimes J'/4 & 0 & 0 & 0
\end{pmatrix}
\]

We can now delete the intermediate rows and columns from these matrices to conclude that

\[
\begin{pmatrix}
2(\sum_{l=1}^{L_1} x_{l} A_{l}^i) \otimes (\sum_{l=1}^{L_1} x_{l'} A_{l'}^i) & 0 \\
0 & 2(\sum_{l=1}^{L_2} y_{l} B_{l}^i) \otimes (\sum_{l=1}^{L_2} y_{l'} B_{l'}^i)
\end{pmatrix}
\succeq
\begin{pmatrix}
0 & J \otimes J'/2 \\
J \otimes J'/2 & 0
\end{pmatrix}
\]

This shows that the constructed points are feasible for \( Q \). Thus, we will have:

\[
\omega(S \otimes S') \leq \omega_{S} \times \omega_{S'}
\]

Combined with our initial observation, this completes the proof. \(\square\)