Parallel Repetition Theorems for Games on Expanders and Related Problems

by

Ricky Rosen

Under the supervision of
Prof. Ran Raz and Prof. Oded Regev

Thesis submitted
for the degree of Doctor of Philosophy

Submitted to the Senate of Tel Aviv University
October 2010
Acknowledgements

First and foremost, I would like to express my deepest gratitude to my advisors, Professor Ran Raz and Professor Oded Regev.

I was fortunate to be a student of Ran Raz. Ran has been a role model for me in every aspect of scientific research. His keen insight and meticulous expertise, his enthusiasm, patience, inspiring ideas, and optimism were precious to me.

I would like to thank Oded Regev for his guidance during my studies and for a great learning experience. His deep intuition, sharp mathematical instincts and high standards were inspiring to me.

I wish to thank Professor Muli Safra for many invaluable advice during my graduate studies. His optimistic way of thinking and his great enthusiasm were an important lesson for me.

During my graduate studies, I met wonderful people who became good friends of mine and a wonderful support group at rough times. I would like to thank Vera Asodi, Greta Yorsh, Svetlana Olonetsky, Iftah Gamzu and Ohad Barzilai.

I would also like to thank my other collaborators on parts of this work, Boaz Barak, Anup Rao and Ronen Shaltiel.

On a more personal note, I would like to thank my parents and my supporting family, my sister Michal, my brother Avi for always being there for me and believing in me.

I would like to thank Golan Pundak for loving me and for making me sometimes believe all those wonderful lies he says about me.
Abstract

Parallel Repetition Theorems for Games on Expanders and Related Problems

Ricky Rosen
Doctor of Philosophy
School of Computer Science
Tel Aviv University

This thesis is concerned with the study of the decrease rate of the value of Multiple Prover Game repeated $n$ times in parallel under different types of models. Our main focus is on Two Prover Game. We also consider multiple-prover games repeated $n$ times in parallel in a special case of a no signaling model.

The parallel repetition theorem states that for any Two Prover Game with value at most $1 - \epsilon$ (for $\epsilon < 1/2$), the value of the game repeated $n$ times in parallel is at most $(1 - \epsilon^3)^\Omega(n/s)$, where $s$ is the length of the answers of the two provers. For Projection Games, the bound on the value of the game repeated $n$ times in parallel was improved to $(1 - \epsilon^2)^\Omega(n)$ and this bound was shown to be tight.

In Chapter 3 we study the case where the underlying distribution, according to which the questions for the two provers are generated, is uniform over the edges of a (bipartite) expander graph. We prove a strong parallel repetition theorem for games of this type.

In Chapter 4 we study the case where the underlying distribution, according to which the questions for the two provers are generated, is a product distribution. This is a special case of expander graph games which are studied in Chapter 3. In this chapter, we give a much simpler proof for this special case and furthermore, the distribution according to which the questions are generated is not necessarily uniform over the edges of a (bipartite) graph.

In Chapter 5 we consider games of $k$-provers where $k > 2$ in a special case of the No-Signaling model. We show that the value of $k$ provers one round game repeated $n$
times in parallel, for $k > 2$, in this special case of the No-Signaling model, decreases exponentially depending only on the value of the original game and on the number of repetitions.
Contents

1 Introduction 1
   1.1 Two-Prover Games 1
   1.2 Parallel Repetition Theorem 2
   1.3 Contribution of the Thesis 3
   1.4 Our Results 7

2 Technical Lemmas 9
   2.1 Preliminaries 9
      2.1.1 Notations 9
      2.1.2 Entropy and Relative Entropy 10
   2.2 Small Relative Entropy and Big Sets 11
   2.3 Main Technical Lemma 13
      2.3.1 Further Notations 13

3 Parallel Repetition Theorems for Games on Expanders 35
   3.1 Introduction 35
      3.1.1 Related Works 36
      3.1.2 Techniques 37
   3.2 Preliminaries 38
      3.2.1 Notations 38
   3.3 Proof of Main Results 43
### 3.3.1 General Games

43

### 3.3.2 Projection Games

50

### 4 Parallel Repetition Theorems for Free Games

63

#### 4.1 Introduction

63

#### 4.2 Preliminaries

65

##### 4.2.1 Notations

65

##### 4.2.2 Special Types of Games

67

#### 4.3 Our Results

67

##### 4.3.1 Technical Lemma

68

##### 4.3.2 Main Lemmas

69

##### 4.3.3 Proof of Main Results

82

### 5 Parallel Repetition Theorems for k-Provers in a No-Signaling Model

85

#### 5.1 Introduction

85

##### 5.1.1 Related Work

88

##### 5.1.2 Our results

89

#### 5.2 Preliminaries

91

#### 5.3 Proof sketch

91

#### 5.4 Technical Lemmas

92

#### 5.5 Parallel Repetition Theorem

95

#### 5.6 Discussion

98

### 6 Discussion and Open Problems

101

#### 6.1 Concluding remarks

101

#### 6.2 Non Product Distribution Between Coordinates

102

#### 6.3 Entangled Provers

103

#### 6.4 More than two player in other models

104

#### 6.5 Unique Games

104
7 Appendix

7.1 Analysis of the Boundaries in the Proof of Lemma (2.3.2) ............... 107
7.2 Proof of Lemma 5.4.1 ................................................. 109
7.3 Proof of Corollary 5.4.2 .............................................. 110
7.4 Generalizing for $k > 3$ .................................................. 110

Bibliography

113
Chapter 1

Introduction

1.1 Two-Prover Games

A two-prover game is played between two players called provers and an additional player called verifier. The game consists of four finite sets $X, Y, A, B$, a probability distribution $P$ over $X \times Y$ and a predicate $V : X \times Y \times A \times B \rightarrow \{0, 1\}$. All parties know $X, Y, A, B, P, V$. The game proceeds as follows. The verifier chooses a pair of questions $(x, y) \in_P X \times Y$ (that is, $(x, y)$ are chosen according to the distribution $P$), and sends $x$ to the first prover and $y$ to the second prover. Each prover knows only the question addressed to her, and the provers are not allowed to communicate with each other. The first prover responds by $a = a(x) \in A$ and the second by $b = b(y) \in B$. The provers jointly win if $V(x, y, a, b) = 1$.

The provers answer the questions according to a pair of functions $a : X \rightarrow A, b : Y \rightarrow B$. The pair $(a, b)$ is called the provers’ strategy or the provers’ protocol. The value of the game is the maximal probability of success that the provers can achieve, where the maximum is taken over all protocols $(a, b)$. That is, the value of the game is

$$\max_{a,b} \mathbb{E}_{(x,y)}[V(x, y, a(x), b(y))]$$

where the expectation is taken with respect to the distribution $P$. 
1.2. PARALLEL REPETITION THEOREM

A two-prover game is called a projection game if for every pair of questions \((x, y) \in X \times Y\) there is a function \(f_{x,y} : B \rightarrow A\), such that, for every \(a \in A\), \(b \in B\), we have: \(V(x, y, a, b) = 1\) if and only if \(f_{x,y}(b) = a\). If in addition, for every \((x, y) \in X \times Y\) the function \(f_{x,y}\) is a bijection (that is, it is one to one and onto), the game is called unique (in which case necessarily \(|A| = |B|\)).

1.2 Parallel Repetition Theorem

Roughly speaking, the parallel repetition of a two-prover game \(G\) is a game where the provers try to win simultaneously \(n\) copies of \(G\). The parallel repetition game is denoted by \(G^\otimes n\). More precisely, in the game \(G^\otimes n\) the verifier generates questions \(x = (x_1, \ldots, x_n) \in X^n\), \(y = (y_1, \ldots, y_n) \in Y^n\), where each pair \((x_i, y_i) \in X \times Y\) is chosen independently according to the original distribution \(P\). The provers respond by \(a = (a_1, \ldots, a_n) = a(x) \in A^n\) and \(b = (b_1, \ldots, b_n) = b(y) \in B^n\). The provers win if they win simultaneously on all \(n\) coordinates, that is, if for every \(i\), we have \(V(x_i, y_i, a_i, b_i) = 1\).

The value of the game \(G^\otimes n\) is not necessarily the same as the value of the game \(G\) raised to the power of \(n\) \([22, 15, 20, 21]\).

The parallel repetition theorem \([32]\) states that for any two-prover game \(G\), with value at most \(1 - \epsilon\) (for any \(0 < \epsilon \leq 1/2\)), the value of the game \(G^\otimes n\) is at most

\[
(1 - \epsilon^c)^{\Omega(n/s)},
\]

(1.1)

where \(s\) is the answers’ length of the original game \((s = |A| \cdot |B|)\), and \(c\) is a universal constant. The constant \(c\) implicit in \([32]\) is \(c = 32\). An example by Feige and Verbitsky \([21]\) shows that the dependency on \(s\) in Equation (1.1) is necessary.

A beautiful recent work by Holenstein \([24]\) simplified the proof of \([32]\) and obtained an improved constant of \(c = 3\). An intriguing followup work by Rao \([31]\) gave for the important special case of projection games, an improved bound of

\[
(1 - \epsilon^2)^{\Omega(n)}.
\]

(1.2)
Several researchers asked whether or not these bounds could be improved to \((1 - \epsilon)^{O(n/s)}\), for general two-prover games, or at least for interesting special cases, such as, projection games, unique games, or xor games (see for example [19, 36]). This question is usually referred to as the strong parallel repetition problem. However, a beautiful analysis shows that the so-called odd cycle game (first studied in [19, 12]) is a counterexample to strong parallel repetition [33]. More precisely, for any \(0 < \epsilon \leq 1/2\), there exists a two-prover game with value \(\leq 1 - \epsilon\), such that, (for large enough \(n\)) the value of the game repeated in parallel \(n\) times is \(\geq (1 - \epsilon^2)^{O(n)}\) [33] (see also [8]). Since the odd cycle game is a projection game, a unique game, and a xor game, this answers negatively most variants of the strong parallel repetition problem. This example also shows that Equation (1.2) is tight.

The parallel repetition theorem was used to prove a large number of hardness of approximation results [9], such as, Håstad’s optimal results on the hardness of approximation of 3-SAT and 3-LIN [23], and Feige’s optimal results on the hardness of approximation of Set-Cover [17]. It also has applications in quantum information theory [12]; in understanding foams and tiling the space \(\mathbb{R}^n\) [19, 27]; and in communication complexity [30, 7]. See [34] for an excellent survey on this subject.

This thesis comprises four main chapters, and in what follows we describe the contents of each of these chapters.

### 1.3 Contribution of the Thesis

In Chapter 2 we prove two technical lemmas.

In Section 2.2 we consider two distributions \(P\) and \(Q\) that are distributed over the same sample space, \(\Omega\), for which the relative entropy of \(P\) with respect to \(Q\) is small. We show that for every set \(S \subseteq \Omega\), for which \(P(S)\) is big enough, \(Q(S)\) cannot be much bigger than \(P(S)\).
1.3. CONTRIBUTION OF THE THESIS

In Section 2.3 we prove our main technical contribution which is a new lemma that may be interesting in its own right. Intuitively, we show that, if two distributions that are distributed over the edges of an expander graph are “close” from the point of view of an average vertex, then they are “close” to each other on the entire space. The proof of the lemma is somewhat involved and uses several new ideas.

References: The results of this chapter appear in:

- A Strong Parallel Repetition Theorem for Projection Games on Expanders,
  Ran Raz, Ricky Rosen. Submitted.

- Strong Parallel Repetition Theorem for Free Projection Games,
  Boaz Barak, Anup Rao, Ran Raz, Ricky Rosen, Ronen Shaltiel.

In Chapter 3 we study the case where the underlying distribution P, according to which the questions for the two provers are generated, is uniform over the edges of a (biregular) bipartite expander graph with sets of vertices \(X, Y\). Let \(M\) be the (normalized) adjacency matrix of the graph and denote by \(1 - \lambda\) the second largest singular value of \(M\). That is, \(\lambda\) is the (normalized) spectral gap of the graph. In this case, we prove that the value of the repeated game is at most

\[
(1 - \epsilon^2)^{\Omega(c(\lambda) \cdot n/s)},
\]

for general games; and at most

\[
(1 - \epsilon)^{\Omega(c(\lambda) \cdot n)},
\]

for projection games, where \(c(\lambda) = \text{poly}(\lambda)\). In particular, for projection games we obtain a strong parallel repetition theorem (when \(\lambda\) is constant).

This gives a strong parallel repetition theorem for a large class of two prover games.
Chapter 1. Introduction

We note that projection games are a general class of games that are used in many applications of the parallel repetition theorem and in particular in most applications for hardness of approximation. We also note that in many applications, and in particular in most applications for hardness of approximation, the underlying graph is a (biregular) bipartite expander graph.

References: The results of this chapter appear in:

- A Strong Parallel Repetition Theorem for Projection Games on Expanders,
  Ran Raz, Ricky Rosen. Submitted.

In Chapter 4 we consider games in which the underlying distribution $P$, according to which the questions for the two provers are generated, is a product distribution. (The term ‘games of no information’ was previously used in [20] to describe such games. The term ‘free’ was introduced in [10] to describe a less general class of games; it was used for games with uniform distribution over full support). The class of free games is a special case of the class of expander games which are studied in Chapter 3. However, in this chapter, we give a much simpler proof for this special case and furthermore, the distribution according to which the questions are generated is not necessarily uniform over the edges of a (bipartite) graph.

We prove that the value of the repeated projection free game is at most

$$(1 - \epsilon)^{\Omega(n)},$$

and for general free games is at most

$$(1 - \epsilon^2)^{\Omega(n/s)}.$$ 

For this case we obtain better constants than in Chapter 3.
1.3. CONTRIBUTION OF THE THESIS

References: The results of this chapter appear in:

- Strong Parallel Repetition Theorem for Free Projection Games,
  Boaz Barak, Anup Rao, Ran Raz, Ricky Rosen, Ronen Shaltiel.

In Chapter 5 we address the problem of the error decrease of parallel repetition games for \( k \)-provers where \( k \geq 2 \). We consider a special case of the No-Signaling model and show that the error of the parallel repetition of \( k \)-provers one-round game, for \( k > 2 \), in this model, decreases exponentially depending only on the error of the original game and on the number of repetitions.

We prove that the value of the repeated no-signaling game is at most

\[
(1 - (\epsilon/k)^2)^{\Omega(n)}.
\]

References: The results of this chapter appear in:

- A \( k \)-Provers Parallel Repetition Theorem for a version of No-Signaling Model, Ricky Rosen.
1.4 Our Results

Recall the definitions from Section 1.3.

We defined $\lambda > 0$ for the case in which the underlying distribution according to which the questions for the two provers are generated, is uniform over the edges of a (biregular) bipartite expander graph with sets of vertices $X, Y$. Let $M$ be the (normalized) adjacency matrix of the graph and denote by $1 - \lambda$ the second largest singular value of $M$. That is, $\lambda$ is the (normalized) spectral gap of the graph.

We now present the main results of this thesis in detail:

**Theorem 1** (Parallel Repetition For General Games). For every game $G$ with value $1 - \epsilon$ where $\epsilon < 1/2$, the value of $G^\otimes n$ is at most $(1 - \epsilon^2 \cdot c(\lambda))^{n/\log s}$ where $s$ is the size of the answers set and $c(\lambda) = (1/32)10^{-12}\lambda^2/(\log(\frac{2}{\lambda}))^2$.

**Theorem 2** (Parallel Repetition For Projection Games). For every projection game $G$ with value $1 - \epsilon$ where $\epsilon < 1/2$, the value of $G^\otimes n$ is at most $(1 - \epsilon)^{\text{poly}(\lambda) \cdot n}$.

**Theorem 3** (Parallel Repetition For Free Games). For every game $G$ with value $1 - \epsilon$ where $\epsilon < 1/2$ and $P_{XY} = P_X \times P_Y$ (the questions are distributed according to some product distribution), the value of $G^\otimes n$ is at most $(1 - \epsilon^2/9)^{n/(18s+3)}$.

**Theorem 4** (Strong Parallel Repetition For Free Projection Games). For every projection game $G$ with value $1 - \epsilon$ where $\epsilon < 1/2$ and $P_{XY} = P_X \times P_Y$ (the questions are distributed according to some product distribution), the value of $G^\otimes n$ is at most $(1 - \epsilon/9)^{(n/33)-1}$.

**Theorem 5** (Parallel Repetition For No-Signaling Games). For every $k \geq 2$ provers game, $G$, with no-signaling value $\omega^{\Diamond}(G) = 1 - \epsilon$ where $\epsilon < 1/2$, the value of $G^\otimes n$ is at most

$$
\left(1 - \frac{\epsilon^2}{100(1 + 4k)^2}\right)^n.
$$
1.4. OUR RESULTS
Chapter 2

Technical Lemmas

In this chapter we prove two technical lemmas that are used in the proofs of the theorems presented in Section 1.4. We begin by introducing notations and basic definitions. We then prove Lemma 2.2 that is used in the proof of Theorem 3 and Theorem 4. Lemma 2.2 is also used for proving Lemma 2.3. Lemma 2.3 is the main technical lemma of this section and is used in the proof of Theorem 1 and Theorem 2. For the proof of Theorem 5, there is no need to use any of the technical lemmas proved in this section.

2.1 Preliminaries

2.1.1 Notations

General Notations

We denote an $n$-dimensional vector by a superscript $n$, e.g., $\phi^n = (\phi_1, \ldots, \phi_n)$ where $\phi_i$ is the $i^{th}$ coordinate. The function $\log(x)$ is the logarithm base 2 of $x$. We use the common notation $[n]$ to denote the set $\{1, \ldots, n\}$.
2.1. PRELIMINARIES

2.1.2 Entropy and Relative Entropy

Definition 2.1.1 (Entropy). For a probability distribution \( \phi \) over a sample space \( \Omega \) we define the entropy of \( \phi \) to be \( H(\phi) = -\sum_{x \in \Omega} \phi(x) \log(\phi(x)) = -\mathbb{E}_{x \sim \phi} \log(\phi(x)) = \mathbb{E}_{x \sim \phi} \log \left( \frac{1}{\phi(x)} \right) \).

By applying Jensen’s inequality on the concave function \( \log(\cdot) \) one can derive the following fact:

Fact 2.1.2. For every distribution \( \phi \) over \( \Omega \), \( H(\phi) \leq \log(|\text{supp}(\phi)|) \) where

\[
\text{supp}(\phi) = \{ x \in \Omega | \phi(x) > 0 \}.
\]

Definition 2.1.3 (Relative Entropy). We define Relative Entropy, also called the Kullback-Leibler Divergence or simply divergence. Let \( P \) and \( Q \) be two probability distributions defined on the same sample space \( \Omega \). The relative entropy of \( P \) with respect to \( Q \) is:

\[
D(P \parallel Q) = \sum_{x \in \Omega} P(x) \log \frac{P(x)}{Q(x)}
\]

where \( 0 \log \frac{0}{0} \) is defined to be \( 0 \) and \( p \log \frac{p}{0} \) where \( p \neq 0 \) is defined to be \( \infty \).

Vaguely speaking, we could think of the relative entropy as a way to measure the information we gained by learning that a random variable is distributed according to \( P \) when a priori we thought that it was distributed according to \( Q \). This indicates how far \( Q \) is from \( P \); if we don’t gain much information then the two distributions are very close in some sense. Note that the relative entropy is not symmetric (and therefore is not a metric).

Fact 2.1.4. Let \( \Phi^n = \Phi_1 \times \Phi_2 \times \cdots \times \Phi_n \) and let \( \mu^n \) be any distribution over the same sample space (not necessarily a product distribution) then \( \sum_{i=1}^n D(\mu_i \parallel \Phi_i) \leq D(\mu^n \parallel \Phi^n) \) thus \( \mathbb{E}_{i \in [n]} D(\mu_i \parallel \Phi_i) = \frac{1}{n} \sum_{i \in [n]} D(\mu_i \parallel \Phi_i) \leq \frac{D(\mu^n \parallel \Phi^n)}{n} \).

For proof see Lemma 3.3 in [32].
2.2 Small Relative Entropy and Big Sets

We show that for all \( p, q \in [0, 1] \), if \( q > 4p \) then \( \text{D}(P\|Q) > 4p \), where \( P = (p, 1 - p) \) and \( Q = (q, 1 - q) \). We will conclude from this lemma that for every two distributions \( P, Q \) (not necessarily a binary distributions) over the same sample space \( \Omega \), if there is a set \( S \in \Omega \), such that \( Q(S) > 4P(S) \) then \( \text{D}(P\|Q) > P(S) \).

**Lemma 2.2.1.** For every \( 0 \leq p, q \leq 1 \) define binary distributions \( P = (p, 1 - p) \) and \( Q = (q, 1 - q) \), over \( \{0, 1\} \). For every \( \delta > p \), if \( \text{D}(P\|Q) \leq \delta \) then

\[
q \leq 4\delta.
\]

**Proof.** If \( \delta \geq \frac{1}{4} \) then the statement is obviously true. For the case that \( \delta < \frac{1}{4} \), assume by way of contradiction that \( q > 4\delta \). Since for \( q > p \), \( \text{D}(P\|Q) \) is decreasing in \( p \) and increasing in \( q \),

\[
\text{D}(P\|Q) = p \log \frac{p}{q} + (1 - p) \log \frac{1 - p}{1 - q} > \delta \log \left( \frac{\delta}{4\delta} \right) + (1 - \delta) \log \frac{1 - \delta}{1 - 4\delta} = -2\delta + (1 - \delta) \log \left( 1 + \frac{3\delta}{1 - 4\delta} \right). \tag{2.1}
\]

If \( \delta \geq 1/7 \) then \( \log \left( 1 + \frac{3\delta}{1 - 4\delta} \right) \geq 1 \). Thus,

\[
(2.1) \geq -2\delta + (1 - \delta) > \delta
\]

where the last inequality follows since \( \delta < 1/4 \).

If \( \delta < 1/7 \) then \( \frac{3\delta}{1 - 4\delta} < 1 \). Using the inequality \( \log_2(1 + x) \geq x \) for every \( 0 \leq x \leq 1 \) we obtain,

\[
(2.1) \geq -2\delta + (1 - \delta) \frac{3\delta}{1 - 4\delta} \geq -2\delta + 3\delta = \delta
\]

where the last inequality follows since \( \frac{1 - \delta}{1 - 4\delta} > 1 \). Since we obtained a contradiction in both cases, the lemma holds.
Corollary 2.2.2. For every probability distributions $P$, $Q$ over the same sample space $\Omega$, for every $T \subseteq \Omega$ and for every $\delta \geq P(T)$, if $D(P\|Q) \leq \delta$ then $Q(T) \leq 4\delta$.

Proof. Denote $p = P(T)$ and $q = Q(T)$ and let $P' = (p, 1 - p)$, $Q' = (q, 1 - q)$. By the data processing inequality for mutual information $D(P\|Q) \geq D(P'\|Q')$ and the corollary follows. \qed
2.3 Main Technical Lemma

2.3.1 Further Notations

Expander Graphs

We will use the notation \((d_X, d_Y)\)-bipartite graph for an unbalanced bipartite regular graph on vertices \(X \cup Y\). That is, a graph where the degree of each vertex \(x \in X\) is \(d_X\), the degree of each vertex \(y \in Y\) is \(d_Y\) and the set of edges of the graph is a subset of \(X \times Y\). In this chapter we work with bipartite expander graphs.

We define a \((X, Y, d_X, d_Y, 1 - \lambda)\)-expander graph \(G_{XY}\) based on the singular values of the normalized adjacency matrix \(M = M(G_{XY})\) of \(G_{XY}\). That is, \(M\) is the adjacency matrix of \(G_{XY}\) where we divide each entry by \(\sqrt{d_X \cdot d_Y}\). We first state a version of the Singular-Value Decomposition Theorem and then explain the definition of \((X, Y, d_X, d_Y, 1 - \lambda)\)-expander graph.

Singular-Value Decomposition Theorem

By the singular-value decomposition theorem, for an \(|X|\)-by-\(|Y|\) matrix \(M\) whose entries come from the field \(\mathbb{R}\), there exists a factorization of the form \(M = U\Sigma V^*\) where \(U\) is an \(|X|\)-by-\(|X|\) unitary matrix, the matrix \(\Sigma\) is \(|X|\)-by-\(|Y|\) diagonal matrix with nonnegative real numbers on the diagonal, and \(V^*\) denotes the conjugate transpose of \(V\), an \(|Y|\)-by-\(|Y|\) unitary matrix. The columns of \(V\) form a set of orthonormal basis vector directions for the rows of \(M\) (these are the eigenvectors of \(M^*M\)). The columns of \(U\) form a set of orthonormal basis vector directions for the columns of \(M\) (these are the eigenvectors of \(MM^*\)). The diagonal values in the matrix \(\Sigma\) are the singular values (these are the square roots of the eigenvalues of \(MM^*\) and \(M^*M\) that correspond with the same columns in \(U\) and \(V\).)

A non-negative real number \(\sigma\) is a singular value for \(M\) if and only if there exist
unit-length vectors $u$ and $v$ such that

$$Mv = \sigma u$$

and

$$M^*u = \sigma v.$$ 

The vectors $u$ and $v$ are called left-singular and right-singular vectors for $\sigma$, respectively.

In any singular value decomposition $M = U\Sigma V^*$ the diagonal entries of $\Sigma$ are equal to the singular values of $M$. The columns of $U$ and $V$ are, respectively, left- and right-singular vectors for the corresponding singular values.

We assume without loss of generality that the singular values are sorted according to their absolute values, that is $\sigma_0 := \Sigma(1,1)$ is the singular value whose absolute value is the largest.

**Singular-Value Decomposition of $M(G_{XY})$**

Because $G_{XY}$ is a $(d_X, d_Y)$-bipartite regular graph and because $M$ is normalized, $\sigma_0 = 1$. Note that all singular values are between 0 and 1. We denote by $1 - \lambda$ the singular value whose value is the closest to 1 and that is not $\sigma_0$. We refer to it as the second singular value. We say that $\lambda$ is the spectral gap of $G_{XY}$.

**Definition of $(X, Y, d_X, d_Y, 1 - \lambda)$-expander graph**

We define $(X, Y, d_X, d_Y, 1 - \lambda)$-expander graph to be a $(d_X, d_Y)$-bipartite graph with second singular value $1 - \lambda$.

It this section distributions $P$ and $Q$ are distributed over the set $X \times Y$. We will use $P_X$ for the marginal distribution of $P$ on $X$, that is, $P_X$ is a distribution on $X$ and

$$P_X(x) = \sum_{y \in Y} P(x, y).$$

For simplicity, we write $P(x)$ rather than $P_X(x)$. Similarly, we use $Q_X, P_Y, Q_Y$ for the marginal distribution of $Q$ on $X$, the marginal distribution of $P$ on $Y$ and the marginal
Chapter 2. Technical Lemmas

distribution of Q on Y respectively. For $y \in Y$, the distribution $P_{X|y}$ is the marginal
distribution of P on $X$ conditioned on $Y = y$, i.e.,

$$P_{X|y}(x) = P(x|y) = P(x, y)/P(y).$$

We use $Q_{X|y}$ for the marginal distribution of Q on $X$ conditioned on $Y = y$. For $x \in X$
we use $P_{Y|x}$ and $Q_{Y|x}$ for the marginal distribution of P on $Y$ conditioned on $X = x$ and
the marginal distribution of Q on $Y$ conditioned on $X = x$ respectively. We will also use
$P_{X|Y}$ for the marginal distribution of P on $X$ conditioned on $Y$ where $Y$ is a random
variable distributed over the set $Y$. For example, we use in Lemma (2.3.1)

$$\mathbb{E}_{y \sim P_Y} D(P_{X|Y} \| Q_{X|Y})$$

this is the expected value of the relative entropy of $P_{X|Y}$ with respect to $Q_{X|Y}$ where
$Y$ is distributed according to $P_Y$. Similarly, $Q_{X|Y}$ for the marginal distribution of Q on
$X$ conditioned on $Y$ where $Y$ is a random variable distributed over the set $Y$, $P_{Y|X}$ for
the marginal distribution of P on $Y$ conditioned on $X$ where $X$ is a random variable
distributed over the set $X$ and $Q_{Y|X}$ for the marginal distribution of Q on $Y$ conditioned
on $X$ where $X$ is a random variable distributed over the set $X$. 
Lemma 2.3.1. Let \( X, Y \) be two sets and let \( d_X, d_Y \) be two integers and \( \lambda > 0 \). Let \( G \) be a \( (X, Y, d_X, d_Y, 1 - \lambda) \)-bipartite expander graph with second singular value \( 1 - \lambda \). Let \( Q \) be the uniform distribution over the edges of \( G \) and let \( P \) be any distribution over the edges of \( G \). For any \( 0 < \epsilon < \alpha < 1/100 \), if the following hold:

\[
\mathbb{E}_{Y \sim P_Y} D(P_{X|Y} \| Q_{X|Y}) \leq \epsilon
\]

and

\[
\mathbb{E}_{X \sim P_X} D(P_{Y|X} \| Q_{Y|X}) \leq \epsilon
\]

then for every \( S \subseteq X \times Y \) such that \( P(S) = \alpha \),

\[
Q(S) < 10^6 \cdot \alpha \cdot \frac{1}{\lambda} \cdot \log \frac{2}{\lambda}.
\]

Intuition  Why do we need the expansion property? Is it true in general that if

\[
\mathbb{E}_{Y \sim P_Y} D(P_{X|Y} \| Q_{X|Y}) \leq \epsilon \quad \text{and} \quad \mathbb{E}_{X \sim P_X} D(P_{Y|X} \| Q_{Y|X}) \leq \epsilon
\]

then for every \( S \subseteq X \times Y \) such that \( P(S) = \alpha \), \( Q(S) \) is not much bigger than \( \alpha \)? We will see a counterexample which appeared at [32]. Let \( P \) and \( Q \) be two distributions over \( \{0, 1\} \times \{0, 1\} \):

\[
P := \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad Q := \begin{pmatrix} \epsilon & 0 \\ 0 & 1 - \epsilon \end{pmatrix}.
\]

Clearly,

\[
\mathbb{E}_{Y \sim P_Y} D(P_{X|Y} \| Q_{X|Y}) = 0 \quad \text{and} \quad \mathbb{E}_{X \sim P_X} D(P_{Y|X} \| Q_{Y|X}) = 0,
\]

and for \( S = \{(1, 1)\} \) although \( P(S) = 0 \), \( Q(S) \) is very big. We can generalize this example. Let \( Q \) be a block diagonal matrix \( A_1 \otimes \ldots \otimes A_m \) such that in each block, \( Q \) is
uniform. Let $P$ be a subset of those blocks such that $P$ is also uniform in each block.

$$P = \begin{pmatrix} A_1 & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \ddots & \cdots & \cdots & \cdots & \vdots \\ 0 & \cdots & A_t & \cdots & \cdots & 0 \\ 0 & 0 & 0 & A_{t+1} & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \cdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & A_m \end{pmatrix}$$

$$Q = \begin{pmatrix} A_1 & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \ddots & \cdots & \cdots & \cdots & \vdots \\ 0 & \cdots & A_t & \cdots & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \cdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Clearly,

$$\mathbb{E}_{Y \sim P_Y} D(P_{X|Y} || Q_{X|Y}) = 0 \text{ and } \mathbb{E}_{X \sim P_X} D(P_{Y|X} || Q_{Y|X}) = 0.$$ 

Let $S$ be any subset of $A_{t+1} \cup \cdots \cup A_m$ and obtain that although $P(S) = 0$, $Q(S)$ could be as big as $1 - \epsilon$. Notice that since we allow

$$\mathbb{E}_{Y \sim P_Y} D(P_{X|Y} || Q_{X|Y}) \text{ and } \mathbb{E}_{X \sim P_X} D(P_{Y|X} || Q_{Y|X})$$

to be at most $\epsilon$ and not strictly 0, there could be edges outside of $A_1, \ldots, A_m$, however, there could be only a small fraction (at most $\epsilon$) of the edges outside of $A_1, \ldots, A_m$.

This case can not happen when $Q$ is a distribution over edges of an expander graph. In that case, there must be many edges between every two sets of vertices who are not too small. Thus, if $Q$ is a uniform distribution over the edges of an expander graph and Equation 2.2 holds, then for every set $S$, whose measure according to $P$ is not too small, $Q(S)$ is not much bigger than $P(S)$.

Proof. of Lemma 2.3.1 For the entire proof fix $X, Y, d_X, d_Y, \lambda, G, Q, S, \epsilon, \alpha$ that satisfy the conditions of the lemma. We denote by $E_G$ the set of edges of $G$ and we denote

$$\alpha' := Q(S).$$

Proof Outline Let $P$ be a distribution over $E_G$. We will show that if $P(S) = \alpha$, and

$$\alpha' := Q(S) \geq 10^6 \cdot \alpha \cdot \frac{1}{\lambda} \cdot \log \frac{2}{\lambda}$$

17
then
\[ \mathbb{E}_{Y \sim P_Y} D(P_{X|Y} \| Q_{X|Y}) + \mathbb{E}_{X \sim P_X} D(P_{Y|X} \| Q_{Y|X}) > 2 \epsilon. \]

To do that we first find a distribution \( P \) that satisfies the conditions above, that is \( P(S) = \alpha \), and minimizes
\[ \mathbb{E}_{Y \sim P_Y} D(P_{X|Y} \| Q_{X|Y}) + \mathbb{E}_{X \sim P_X} D(P_{Y|X} \| Q_{Y|X}). \]

This is done in Lemma (2.3.2). We then study this distribution and obtain bounds on its values, i.e., bounds on \( P(x, y) \) for all \( x, y \), (this is done in Lemma (2.3.3)). Later we show that this distribution is close to the uniform distribution over the edges of \( G \) (Lemma (2.3.4)). For this distribution \( P \), we obtain a contradiction.

**Lemma 2.3.2.** Let \( P \) be a distribution over \( E_G \) that minimizes
\[ \mathbb{E}_{Y \sim P_Y} D(P_{X|Y} \| Q_{X|Y}) + \mathbb{E}_{X \sim P_X} D(P_{Y|X} \| Q_{Y|X}) \]
under the constraint \( P(S) = \alpha \). Then there exist constants \( c_0, c_1 > 0 \) such that for every \((x, y) \in E_G:\)
\[
P(x, y) = \begin{cases} 
c_1 \sqrt{P(x)} \cdot \sqrt{P(y)} & (x, y) \in S; \\
c_0 \sqrt{P(x)} \cdot \sqrt{P(y)} & (x, y) \notin S.
\end{cases}
\]

**Proof.** Since \( Q \) is a uniform distribution over a \((d_X, d_Y)\)--bipartite graph, for all \((x, y) \in E_G \subseteq X \times Y, Q(y|x) = 1/d_X \) and \( Q(x|y) = 1/d_Y \). By definition,
\[
\mathbb{E}_{X \sim P_X} D(P_{Y|X} \| Q_{Y|X}) = \sum_{(x, y) \in X \times Y} P(x, y) \log \left( \frac{P(x, y)}{P(x)} \right) \frac{Q(x)}{Q(y)} \log \left( \frac{Q(x)}{Q(y)} \right)
\]
\[
= \sum_{(x, y) \in X \times Y} P(x, y) \log P(x, y) - \sum_{x \in X} P(x) \log P(x) - \log(1/d_X)
\]
and similarly,
\[
\mathbb{E}_{Y \sim P_Y} D(P_{X|Y} \| Q_{X|Y}) = \sum_{(x, y) \in X \times Y} P(x, y) \log \left( \frac{P(x, y)}{P(y)} \right) \frac{Q(y)}{Q(x)} \log \left( \frac{Q(y)}{Q(x)} \right)
\]
\[
= \sum_{(x, y) \in X \times Y} P(x, y) \log P(x, y) - \sum_{y \in Y} P(y) \log P(y) - \log(1/d_Y).\]
Thus,
\[
\mathbb{E}_{X \sim P_X} D(P_{Y|X} \| Q_{Y|X}) + \mathbb{E}_{Y \sim P_Y} D(P_{X|Y} \| Q_{X|Y}) = \\
2 \sum_{(x,y) \in X \times Y} P(x,y) \log P(x,y) - \sum_{x \in X} P(x) \log P(x) - \sum_{y \in Y} P(y) \log P(y) \\
- \log(1/d_X) - \log(1/d_Y).
\]
(2.3)

We will minimize Equation (2.3) under the constraints:
\[
\sum_{(x,y) \in S} P(x,y) = \alpha \\
\sum_{(x,y) \notin S} P(x,y) = 1 - \alpha
\]
where $1 \geq P(x,y) \geq 0$. We will use Lagrange multipliers to find the minimum\(^1\) of Equation (2.3). We define the Lagrange function by
\[
\Lambda = 2 \sum_{(x,y) \in X \times Y} P(x,y) \log P(x,y) - \sum_{x \in X} P(x) \log P(x) - \sum_{y \in Y} P(y) \log P(y) - \log(1/d_X) \\
- \log(1/d_Y) - \lambda_1 \left( \sum_{(x,y) \in S} P(x,y) - \alpha \right) - \lambda_2 \left( \sum_{(x,y) \notin S} P(x,y) - 1 + \alpha \right).
\]

For every $(x,y) \in S$:
\[
\frac{\partial \Lambda}{\partial P(x,y)} = 2 \log P(x,y) - \log P(x) - \log P(y) - \lambda_1 = 0.
\]

For every $(x,y) \notin S$:
\[
\frac{\partial \Lambda}{\partial P(x,y)} = 2 \log P(x,y) - \log P(x) - \log P(y) - \lambda_2 = 0.
\]

Thus for every $(x,y) \in S$:
\[
\log \left( \frac{\frac{P(x,y)^2}{P(x) \cdot P(y)}}{\sqrt{P(x) \cdot P(y)}} \right) = \lambda_1.
\]

\(^1\)We will use Lagrange multipliers to find the minimum for $1 > P(x,y) > 0$. In Appendix 7.1 we deal with $P(x,y) = 0$ and $P(x,y) = 1$. 

19
For every \((x, y) \notin S\) :
\[
\log \left( \frac{P(x, y)^2}{P(x) \cdot P(y)} \right) = \lambda_2.
\]
Fixing \(c_0 := 2^{(1/2) \cdot \lambda_2}\) and \(c_1 := 2^{(1/2) \cdot \lambda_1}\) we obtain:
\[
P(x, y) = \begin{cases} 
    c_1 \sqrt{P(x)} \cdot \sqrt{P(y)} & (x, y) \in S; \\
    c_0 \sqrt{P(x)} \cdot \sqrt{P(y)} & (x, y) \notin S.
\end{cases}
\]

For the rest of the proof, we fix \(P\) to be the distribution as in Lemma (2.3.2), that is, \(P\) is the distribution over \(E_G\) that minimizes
\[
\mathbb{E}_{Y \sim P_Y} D(P_{X|Y} \| Q_{X|Y}) + \mathbb{E}_{X \sim P_X} D(P_{Y|X} \| Q_{Y|X})
\]
under the constraint \(P(S) = \alpha\), and let \(c_0, c_1\) be the constants such that for every \((x, y) \in E_G:\n\[
P(x, y) = \begin{cases} 
    c_1 \sqrt{P(x)} \cdot \sqrt{P(y)} & (x, y) \in S; \\
    c_0 \sqrt{P(x)} \cdot \sqrt{P(y)} & (x, y) \notin S.
\end{cases}
\]
For the entire proof we denote by \(\delta\) the smallest non-negative real number for which
\[
\sum_{(x,y) \in S} P(x) - Q(x) \leq \delta \cdot d_X
\]
\[
\sum_{(x,y) \in S} P(y) - Q(y) \leq \delta \cdot d_Y.
\]

**Lemma 2.3.3.** For every \(\delta \geq 0\), if the following hold:
\[
\sum_{(x,y) \in S} P(x) - Q(x) \leq \delta \cdot d_X
\]
\[
\sum_{(x,y) \in S} P(y) - Q(y) \leq \delta \cdot d_Y
\]
then\(^2\)
\[
\frac{1 - \alpha}{\sqrt{d_X \cdot d_Y}} \leq c_0 \leq \frac{1 + 5 \alpha \log ((\alpha' + \delta)/\alpha)}{\sqrt{d_X \cdot d_Y}}
\]
\(^2\)We will only use the upper bound on \(c_0\).
and
\[
\frac{\alpha/(\alpha' + \delta)}{\sqrt{d_X \cdot d_Y}} \leq c_1 \leq \frac{8}{\sqrt{d_X \cdot d_Y}}
\]
(where \(\alpha' := Q(S)\)).

**Proof.** First we will bound \(c_0\) from below. By Cauchy-Schwarz inequality,
\[
(1 - \alpha)^2 = \left( \sum_{(x,y) \notin S} P(x,y) \right)^2 = c_0^2 \cdot \left( \sum_{(x,y) \notin S} \sqrt{P(x)} \cdot \sqrt{P(y)} \right)^2
\]
\[
\leq c_0^2 \cdot \left( \sum_{(x,y) \notin S} P(x) \cdot \sum_{(x,y) \notin S} P(y) \right).
\]
(2.4)

Since for every \(x \in X\) there are at most \(d_X\) elements \(y \in Y\) for which \((x, y) \notin S\) and for every \(y \in Y\) there are at most \(d_Y\) elements \(x \in X\) for which \((x, y) \notin S\), we can bound Equation (2.4) by \(c_0^2 \cdot (d_X \cdot d_Y)\). Therefore,
\[
c_0 \geq \frac{1 - \alpha}{\sqrt{d_X \cdot d_Y}}
\]
(2.5)

Next we will bound \(c_1\) from below.
\[
\alpha^2 = \left( \sum_{(x,y) \in S} P(x,y) \right)^2 = c_1^2 \cdot \left( \sum_{(x,y) \in S} \sqrt{P(x)} \cdot \sqrt{P(y)} \right)^2
\]
\[
\leq c_1^2 \cdot \left( \sum_{(x,y) \in S} P(x) \cdot \sum_{(x,y) \in S} P(y) \right).
\]
(2.6)

Since
\[
\sum_{(x,y) \in S} P(x) = \sum_{(x,y) \in S} Q(x) + \sum_{(x,y) \in S} P(x) - Q(x) \leq d_X (\alpha' + \delta)
\]
\[
\sum_{(x,y) \in S} P(y) = \sum_{(x,y) \in S} Q(y) + \sum_{(x,y) \in S} P(y) - Q(y) \leq d_Y (\alpha' + \delta)
\]
we obtain:
\[
\alpha^2 \leq c_1^2 \cdot d_X d_Y (\alpha' + \delta)^2.
\]
2.3. MAIN TECHNICAL LEMMA

Thus,

\[ c_1 \geq \frac{\alpha/(\alpha' + \delta)}{\sqrt{d_X \cdot d_Y}} \quad (2.7) \]

We now prove upper bounds on \( c_0, c_1 \). Using

\[ \mathbb{E}_{Y \sim P_Y} D(P_{X|Y} || Q_{X|Y}) + \mathbb{E}_{X \sim P_X} D(P_{Y|X} || Q_{Y|X}) \leq 2\epsilon \]

we obtain:

\[
2\epsilon \geq \sum_{(x,y) \in X \times Y} P(x,y) \log \left( \frac{P(x,y)/P(x)}{Q(x,y)/Q(x)} \right) + \sum_{(x,y) \in X \times Y} P(x,y) \log \left( \frac{P(x,y)/P(y)}{Q(x,y)/Q(y)} \right)
\]

\[
= \sum_{(x,y) \in S} P(x,y) \log \left( \frac{P(x,y)/P(x)}{Q(x,y)/Q(x)} \right) + \sum_{(x,y) \notin S} P(x,y) \log \left( \frac{P(x,y)/P(y)}{Q(x,y)/Q(y)} \right) + \sum_{(x,y) \notin S} P(x,y) \log \left( \frac{P(x,y)/P(y)}{Q(x,y)/Q(y)} \right)
\]

\[
= \sum_{(x,y) \in S} P(x,y) \log \left( \frac{c_1 \sqrt{P(y)/P(x)}}{1/d_X} \right) + \sum_{(x,y) \notin S} P(x,y) \log \left( \frac{c_0 \sqrt{P(y)/P(x)}}{1/d_X} \right)
\]

\[
= \sum_{(x,y) \in S} P(x,y) \log \left( \frac{c_1 \sqrt{P(y)/P(x)}}{1/d_X} \cdot \frac{c_1 \sqrt{P(y)/P(x)}}{1/d_X} \right) + \sum_{(x,y) \notin S} P(x,y) \log \left( \frac{c_0 \sqrt{P(y)/P(x)}}{1/d_X} \cdot \frac{c_0 \sqrt{P(y)/P(x)}}{1/d_X} \right)
\]

\[
= \sum_{(x,y) \in S} P(x,y) \log \left( \frac{c_1^2}{1/d_X \cdot 1/d_Y} \right) + \sum_{(x,y) \notin S} P(x,y) \log \left( \frac{c_0^2}{1/d_X \cdot 1/d_Y} \right)
\]

\[
= \alpha \log \left( c_1^2 d_X d_Y \right) + (1-\alpha) \log \left( c_0^2 d_X d_Y \right).
\]

Where the last equality follows from \( \sum_{(x,y) \in S} P(x,y) = \alpha \) and \( \sum_{(x,y) \notin S} P(x,y) = 1 - \alpha. \)
Using Equation (2.5) we obtain:

\[ 2\epsilon \geq \alpha \log (c_1^2 \cdot d_X \cdot d_Y) + (1 - \alpha) \log ((1 - \alpha)^2) \geq \alpha \log (c_1^2 \cdot d_X \cdot d_Y) - 3\alpha \]

Thus,

\[ \log (c_1^2 \cdot d_X \cdot d_Y) \leq 2\epsilon / \alpha + 3 \leq 5 \]

Therefore,

\[ c_1 \leq \frac{8}{\sqrt{d_X \cdot d_Y}} \]

Using Equation (2.7) we obtain:

\[ 2\epsilon \geq \alpha \log (\alpha^2 / (\alpha' + \delta)^2) + (1 - \alpha) \log (c_0^2 \cdot d_X \cdot d_Y) \]

since \( \epsilon < \alpha < 1/100 \) and since \( \alpha' > 2\alpha \) and since \( 2\epsilon + 2\alpha \log ((\alpha' + \delta) / \alpha) \leq 1/2 \) (because \( \alpha' + \delta \leq 2 \) and \( \epsilon, \alpha < 1/100 \)),

\[ \log (c_0 \cdot \sqrt{d_X \cdot d_Y}) \leq 2\epsilon + 2\alpha \log ((\alpha' + \delta) / \alpha) \leq \log (1 + 4\epsilon + 4\alpha \log ((\alpha' + \delta) / \alpha)) \]

\[ \leq \log (1 + 5\alpha \log ((\alpha' + \delta) / \alpha)) \]

Thus,

\[ c_0 \leq (1 + 5\alpha \log ((\alpha' + \delta) / \alpha)) / \sqrt{d_X \cdot d_Y} \]

We now prove that P is close to the uniform distribution over \( E_G \).

**Lemma 2.3.4.** The following hold:

1. \( \sum_{x \in X} \frac{1}{\sqrt{|X|}} \cdot \sqrt{P(x)} \geq 1 - 16\alpha \log (\alpha' / \alpha) / (2\lambda - \lambda^2) \)

2. \( \sum_{y \in Y} \frac{1}{\sqrt{|Y|}} \cdot \sqrt{P(y)} \geq 1 - 16\alpha \log (\alpha' / \alpha) / (2\lambda - \lambda^2) \)
2.3. MAIN TECHNICAL LEMMA

Proof. Define $M$ to be $|X| \times |Y|$ matrix where

$$M_{x,y} = \begin{cases} 
\frac{1}{\sqrt{d_x \cdot d_y}} & \text{if } (x, y) \in E_G; \\
0 & \text{otherwise.}
\end{cases}$$

Define $M'$ to be $|X| \times |Y|$ matrix where

$$M'_{x,y} = \begin{cases} 
c_1 & (x, y) \in S \text{ and } (x, y) \in E_G; \\
c_0 & (x, y) \notin S \text{ and } (x, y) \in E_G; \\
0 & \text{otherwise.}
\end{cases}$$

Recall that

$$c_1 = \frac{P(x, y)}{\sqrt{P(x)P(y)}} \quad \text{for } (x, y) \in S \text{ and } (x, y) \in E_G; \quad (2.9)$$

$$c_0 = \frac{P(x, y)}{\sqrt{P(x)P(y)}} \quad \text{for } (x, y) \notin S \text{ and } (x, y) \in E_G.$$

Denote by $\sqrt{P(X)}$ and $\sqrt{P(Y)}$ the vectors

$$\left(\sqrt{P(x_1)}, \ldots, \sqrt{P(x_{|X|})}\right)$$

and

$$\left(\sqrt{P(y_1)}, \ldots, \sqrt{P(y_{|Y|})}\right)$$

respectively.

Claim 2.3.5. $\sqrt{P(X)} \cdot M' = \sqrt{P(Y)}$

Proof.

$$\left(\sqrt{P(X)} \cdot M'\right)_i = \left(\sqrt{P(x_1)}, \ldots, \sqrt{P(x_{|X|})}\right) \begin{pmatrix} 
P(x_1, y_i)/\sqrt{P(x_1)P(y_i)} \\
\vdots \\
P(x_{|X|}, y_i)/\sqrt{P(x_{|X|})P(y_i)}
\end{pmatrix}$$

$$= \sum_{j=1}^{|X|} \frac{P(x_j, y_i)}{\sqrt{P(y_i)}} = \sqrt{P(y_i)}$$
Claim 2.3.6. $\| \sqrt{P(X)} \cdot M \|_2 > 1 - 8\alpha \log ((\alpha' + \delta)/\alpha)$

Proof. Define $\tilde{M} := M' - M$. That is, $\tilde{M}$ is the following $|X| \times |Y|$ matrix:

$$\tilde{M}_{x,y} = \begin{cases} c_1(1 - 1/(c_1 \cdot \sqrt{d_X \cdot d_Y})) & (x, y) \in S \text{ and } (x, y) \in E_G; \\ c_0(1 - 1/(c_0 \cdot \sqrt{d_X \cdot d_Y})) & (x, y) \notin S \text{ and } (x, y) \in E_G; \\ 0 & \text{else.} \end{cases}$$

By Cauchy-Schwarz inequality and since $\| \sqrt{P(Y)^T} \|_2 = 1$,

$$\| \sqrt{P(X)} \cdot M \|_2 = \| \sqrt{P(X)} \cdot M \|_2 \cdot \| \sqrt{P(Y)^T} \|_2 \geq \sqrt{P(X)} \cdot M \cdot \sqrt{P(Y)^T}$$

$$= \sqrt{P(X)} \cdot M' \cdot \sqrt{P(Y)^T} - \sqrt{P(X)} \cdot \tilde{M} \sqrt{P(Y)^T}$$

Using Claim (2.3.5) we obtain

$$\sqrt{P(X)} \cdot M' \cdot \sqrt{P(Y)^T} = 1$$

Therefore, using Equation (2.9) we obtain,

$$\| \sqrt{P(X)} \cdot M \|_2 \geq 1 - \sqrt{P(X)} \cdot \tilde{M} \sqrt{P(Y)^T}$$

$$= 1 - \alpha \left( 1 - 1/(c_1 \cdot \sqrt{d_X \cdot d_Y}) \right) - (1 - \alpha)(1 - 1/(c_0 \cdot \sqrt{d_X \cdot d_Y}))$$

$$\geq 1 - \alpha - (1 - \alpha)(1 - 1/(c_0 \cdot \sqrt{d_X \cdot d_Y}))$$

$$= (1 - \alpha)/(c_0 \cdot \sqrt{d_X \cdot d_Y})$$

By the bound we have on $c_0$ in Lemma (2.3.3)

$$\| \sqrt{P(X)} \cdot M \|_2 \geq (1 - \alpha)/(1 + 5\alpha \log ((\alpha' + \delta)/\alpha)) \geq 1 - 8\alpha \log ((\alpha' + \delta)/\alpha)$$

Let $U \Sigma V^*$ be the singular value decomposition of $M$ and let $\{u_1, \ldots, u_{|X|}\}$ be the rows of $U$ and recall that this set is an orthonormal basis. Recall that $1 - \lambda$ is the second singular value of $M$. Let $\sqrt{P(X)} = \sum_{i=1}^{|X|} a_i u_i$, that is, $\sqrt{P(X)}$ represented according
to the orthonormal basis \( \{u_1, \ldots, u_{|X|}\} \). Since \( \{u_1, \ldots, u_{|X|}\} \) is an orthonormal basis, \( \sum_{i=1}^{|X|} a_i^2 = 1 \) and also:

\[
\|\sqrt{P(X)} \cdot M\|_2^2 \leq a_1^2 + (1 - \lambda)^2 \sum_{i=2}^{|X|} a_i^2 \\
= (1 - (1 - \lambda)^2) a_1^2 + (1 - \lambda)^2 \sum_{i=1}^{|X|} a_i^2 \\
= (1 - (1 - \lambda)^2) a_1^2 + (1 - \lambda)^2 = (2\lambda - \lambda^2) a_1^2 + (1 - \lambda)^2
\]

Thus,

\[
a_1^2 (2\lambda - \lambda^2) \geq \|\sqrt{P(X)} \cdot M\|_2^2 - (1 - \lambda)^2
\]

We now obtain,

\[
a_1 \geq a_1^2 \geq \|\sqrt{P(X)} \cdot M\|_2^2/(2\lambda - \lambda^2) - 1/(2\lambda - \lambda^2) + 1 \\
\geq (1 - 8\alpha \log ((\alpha' + \delta)/\alpha)^2)/(2\lambda - \lambda^2) - 1/(2\lambda - \lambda^2) + 1 \\
\geq 1 - 16\alpha \log ((\alpha' + \delta)/\alpha)/(2\lambda - \lambda^2)
\]

Since \( a_1 \) is the inner product of the vector \( \sqrt{P(X)} \) with the vector \( u_1 = \left( \frac{1}{\sqrt{|X|}}, \ldots, \frac{1}{\sqrt{|X|}} \right) \),

\[
a_1 = \sum_{x \in X} \frac{1}{\sqrt{|X|}} \sqrt{P(x)} \geq 1 - 16\alpha \log ((\alpha' + \delta)/\alpha)/(2\lambda - \lambda^2)
\]

In the same way we can derive

\[
\sum_{y \in Y} \frac{1}{\sqrt{|Y|}} \sqrt{P(y)} \geq 1 - 16\alpha \log ((\alpha' + \delta)/\alpha)/(2\lambda - \lambda^2)
\]

\[\square\]

\textbf{Claim 2.3.7.} \( \delta \leq 16\alpha' \)

\textit{Proof.} Without loss of generality assume that

\[
\frac{1}{d_X} \cdot \sum_{(x,y) \in S} P(x) - Q(x) \geq \frac{1}{d_Y} \cdot \sum_{(x,y) \in S} P(y) - Q(y).
\]
Hence by the definition of $\delta$, if $\delta > 0$ we have, $\delta \cdot d_X = \sum_{(x,y) \in S} P(x) - Q(x)$. We can bound $\sum_{(x,y) \in S} P(x) - Q(x)$ by $d_X \sum_{x \in S'} P(x) - Q(x)$ where $S' \subseteq X$ is the set of size $\alpha' |X|$ with the largest values of $P(x)$.

Recall that

$$\sum_{x \in X} \frac{1}{\sqrt{|X|}} \sqrt{P(x)} \geq 1 - 16\alpha \log \left( (\alpha' + \delta)/\alpha \right)/(2\lambda - \lambda^2). \quad (2.10)$$

We will assume by way of contradiction that $\delta > 16\alpha'$ and obtain a contradiction to Equation (2.10). Since $\sqrt{\cdot}$ is a concave function, we can assume (for the proof of this claim only) that the distribution $P_X$ is uniform inside $S'$ and uniform outside of $S'$ (by redistributing the mass inside $S'$ and redistributing the mass outside $S'$). Since $\sum_{x \in S'} P(x) - Q(x) \geq \delta$, we can assume in order to derive a contradiction to Equation (2.10) that $\sum_{x \in S'} P(x) - Q(x) = \delta$. Thus we can assume that $P_X$ is as follows:

$$P(x) = \begin{cases} \frac{1}{|X|} (1 + \delta/\alpha') & x \in S' ; \\ \frac{1}{|X|} (1 - \delta/(1 - \alpha')) & x \notin S'. \end{cases}$$

Thus

$$\sum_{x \in X} \frac{1}{\sqrt{|X|}} \sqrt{P(x)} = \alpha' \cdot \sqrt{1 + \delta/\alpha'} + (1 - \alpha') \sqrt{1 - \delta/(1 - \alpha')}$$

$$\leq \alpha' \cdot \left( 1 + \sqrt{\delta/\alpha'} \right) + (1 - \alpha') \left( 1 - \delta/(2(1 - \alpha')) \right)$$

$$= \sqrt{\delta} \cdot \alpha' + 1 - \delta/2$$

Since $\delta > 16\alpha'$, $\sqrt{\delta} \cdot \alpha' + 1 - \delta/2 \leq 1 - \delta/4$. Hence by Equation (2.10)

$$\delta/4 < 16\alpha \log \left( (\alpha' + \delta)/\alpha \right)/(2\lambda - \lambda^2)$$

Since $0 \leq \lambda \leq 1$

$$\delta/4 < 16\alpha \log ((\alpha' + \delta)/\alpha)/\lambda$$

Since $\alpha' < (1/16)\delta$

$$\delta/\alpha < 64 \log ((17/16)\delta/\alpha)/\lambda$$
2.3. MAIN TECHNICAL LEMMA

That is

\[ \frac{\delta}{\alpha} - 64 \log \left( \frac{17}{16} \frac{\delta}{\alpha} \right) / \lambda < 0 \]  \hspace{1cm} (2.11) 

Since \( \delta \geq 16\alpha' \geq 16 \cdot 10^6 \cdot \alpha \cdot \frac{1}{\lambda} \cdot \log \frac{2}{\lambda} \) we have that \( \frac{\delta}{\alpha} \geq 10000 \cdot \frac{1}{\lambda} \cdot \log \frac{2}{\lambda} \). Note that for \( \frac{\delta}{\alpha} = 10000 \cdot \frac{1}{\lambda} \cdot \log \frac{2}{\lambda} \) we obtain a contradiction. To see this observe that since \( \log \frac{2}{\lambda} > 1 \), we can derive:

\[
\frac{\delta}{\alpha} - 64 \log \left( \frac{17}{16} \frac{\delta}{\alpha} \right) / \lambda = 10000 \cdot \frac{1}{\lambda} \cdot \log \frac{2}{\lambda} - \frac{64}{\lambda} \log \left( \frac{17}{16} \cdot 10000 \cdot \frac{1}{\lambda} \cdot \log \frac{2}{\lambda} \right) 
\]
\[ = 10000 \cdot \frac{1}{\lambda} \cdot \log \frac{2}{\lambda} - 64 \log 5312.5 - \frac{64}{\lambda} \log \frac{2}{\lambda} - \frac{64}{\lambda} \log \log \frac{2}{\lambda} 
\]
\[ \geq 9936 \cdot \frac{1}{\lambda} \cdot \log \frac{2}{\lambda} - \frac{64}{\lambda} \log 5312.5 - \frac{64}{\lambda} \log \frac{2}{\lambda} 
\]
\[ \geq 9872 \cdot \frac{1}{\lambda} \cdot \log \frac{2}{\lambda} - 832 \geq 9872 \cdot \frac{1}{\lambda} - 832 = 9040 \cdot \frac{1}{\lambda} > 0 
\]

To obtain a contradiction for all \( \frac{\delta}{\alpha} > 10000 \cdot \frac{1}{\lambda} \cdot \log \frac{2}{\lambda} \) we now show that the derivative of the function

\[ f(z) = z - 64 \log \left( \frac{17}{16} z \right) / \lambda \]

is positive for \( z > 10000 \cdot \frac{1}{\lambda} \cdot \log \frac{2}{\lambda} \) and hence, \( f(z) > 0 \) for all

\[ z \geq 10000 \cdot \frac{1}{\lambda} \cdot \log \frac{2}{\lambda} \].

Substituting \( z = \delta/\alpha \) we obtain a contradiction to Equation (2.11).

\[ \frac{df}{dz} = 1 - \frac{64}{\lambda \ln 2} \cdot \frac{1}{z} > 0 \]

where the last inequality follows since \( z > 64 / (\lambda \ln 2) \)

Corollary 2.3.8.

\[
\sum_{x \in X} \frac{1}{\sqrt{|X|}} \sqrt{P(x)} \geq 1 - 16(\alpha/\lambda) \log (17\alpha'/\alpha) 
\]
\[
\sum_{y \in Y} \frac{1}{\sqrt{|Y|}} \sqrt{P(y)} \geq 1 - 16(\alpha/\lambda) \log (17\alpha'/\alpha) 
\]
Chapter 2. Technical Lemmas

Proof. By Lemma (2.3.4)

1.

\[ \sum_{x \in X} \frac{1}{\sqrt{|X|}} \cdot \sqrt{P(x)} \geq 1 - 16\alpha \log \left( \frac{(\alpha' + \delta)/\alpha}{2\lambda - \lambda^2} \right) \]

\[ = 1 - 16(\alpha/\lambda) \log \left( \frac{(\alpha' + \delta)/\alpha}{2 - \lambda} \right) \]

2.

\[ \sum_{y \in Y} \frac{1}{\sqrt{|Y|}} \cdot \sqrt{P(y)} \geq 1 - 16\alpha \log \left( \frac{(\alpha' + \delta)/\alpha}{2\lambda - \lambda^2} \right) \]

\[ = 1 - 16(\alpha/\lambda) \log \left( \frac{(\alpha' + \delta)/\alpha}{2 - \lambda} \right) \]

The corollary follows by the bound on \( \delta \) in Claim (2.3.7), (\( \delta \leq 16\alpha' \)) and since

\[ 1 \leq 2 - \delta \leq 2. \]

\[ \square \]

We denote

\[ X_1 = \{ x \in X \mid P(x)/Q(x) < 1/2 \} \]
\[ X_2 = \{ x \in X \mid 1/2 \leq P(x)/Q(x) \leq 2 \} \]
\[ X_3 = \{ x \in X \mid P(x)/Q(x) > 2 \} \]

For the set \( S \) given in the lemma denote

\[ S_1 = \{ (x, y) \in S \mid x \in X_1 \} \]
\[ S_2 = \{ (x, y) \in S \mid x \in X_2 \} \]
\[ S_3 = \{ (x, y) \in S \mid x \in X_3 \} \]

We will show that \( Q(S_1), Q(S_3) \) are small since \( Q(X_1) \) and \( Q(X_3) \) are small and that \( Q(S_2) \leq 40\alpha \)
Claim 2.3.9. $Q(S_1) \leq 400(\alpha/\lambda) \log (17\alpha'/\alpha)$

Proof. Note that $Q(S_1) \leq Q(X_1)$. We will assume by way of contradiction that

$$Q(X_1) > 400(\alpha/\lambda) \log (17\alpha'/\alpha)$$

and obtain a contradiction to Corollary (2.3.8). Since $\sqrt{\cdot}$ is a concave function, we can assume (for the proof of this claim only) that the distribution $P_X$ is uniform inside $X_1$ and uniform outside of $X_1$ (by redistributing the mass inside $X_1$ and redistributing the mass outside $X_1$), (note that this doesn’t change the definition of $X_1$). Since for every $x \in X_1 \ P(x)/Q(x) < 1/2$, we can assume in order to derive a contradiction that for every $x \in X_1 \ P(x)/Q(x) = 1/2$. Denote

$$q := Q(X_1).$$

Thus we can assume that $P_X$ is as follows:

$$P(x) = \begin{cases} 
\frac{1}{|X|} & x \in X_1; \\
\frac{1}{|X|} \left(1 + \frac{q}{2(1-q)} \right) & x \notin X_1.
\end{cases}$$

Thus

$$\sum_{x \in X} \frac{1}{|X|} \sqrt{P(x)} = q \cdot 1/\sqrt{2} + (1-q) \sqrt{1 + \frac{q}{2(1-q)}}$$

$$\leq q/\sqrt{2} + (1-q)(1 + \frac{q}{4(1-q)})$$

$$= q/\sqrt{2} + (1-q) + \frac{q}{4}$$

$$< 1 - 0.04q$$

Hence by Corollary (2.3.8)

$$0.04q < 16(\alpha/\lambda) \log (17\alpha'/\alpha).$$

Thus,

$$q < 400(\alpha/\lambda) \log (17\alpha'/\alpha).$$
Claim 2.3.10. $Q(S_3) \leq 200(\alpha/\lambda) \log (17\alpha'/\alpha)$

Proof. Note that

$$Q(S_3) \leq Q(X_3).$$

We will assume by way of contradiction that $Q(X_3) > 200(\alpha/\lambda) \log (17\alpha'/\alpha)$ and obtain a contradiction to Corollary (2.3.8). Since $\sqrt{\cdot}$ is a concave function, we can assume (for the proof of this claim only) that the distribution $P_X$ is uniform inside $X_3$ and uniform outside of $X_3$ (by redistributing the mass inside $X_3$ and redistributing the mass outside $X_3$), (note that this doesn’t change the definition of $X_3$). Since for every $x \in X_3$ $P(x)/Q(x) > 2$, we can assume in order to derive a contradiction that for every $x \in X_3$ $P(x)/Q(x) = 2$. Denote

$$q := Q(X_3).$$

Thus we can assume that $P_X$ is as follows:

$$P(x) = \begin{cases} \frac{2}{|X|} & x \in X_3 ; \\ \frac{1}{|X|} \left(1 - \frac{q}{1-q}\right) & x \notin X_3. \end{cases}$$

Thus

$$\sum_{x \in X} \frac{1}{\sqrt{|X|}} \sqrt{P(x)} = q \cdot \sqrt{2} + (1 - q) \sqrt{1 - \frac{q}{1-q}}$$

$$\leq q\sqrt{2} + (1 - q)(1 - \frac{q}{2(1-q)})$$

$$= q\sqrt{2} + (1 - q) - \frac{q}{2}$$

$$< 1 - 0.08q$$

Hence by Corollary (2.3.8)

$$0.08q < 16(\alpha/\lambda) \log (17\alpha'/\alpha).$$

Thus,

$$q < 200(\alpha/\lambda) \log (17\alpha'/\alpha).$$
2.3. MAIN TECHNICAL LEMMA

Claim 2.3.11. \( Q(S_2) \leq 40\alpha \)

Proof. Denote

\[
\tilde{P}(x, y) = \begin{cases} 
Q(x)P(y|x) & x \in X_2; \\
Q(x, y) & \text{otherwise.}
\end{cases}
\]

The proof will follow by combining Claim (2.3.12) and Claim (2.3.13). We will show that \( Q(S_2) \leq 4\tilde{P}(S_2) + 16\alpha \) and \( \tilde{P}(S_2) \leq 2P(S_2) \leq 2\alpha. \)

Claim 2.3.12. \( \tilde{P}(S_2) \leq 2P(S_2) \)

Proof. Since for every \( x \in X_2 \) \( Q(x) \leq 2P(x) \)

\[
\tilde{P}(S_2) = \sum_{(x,y) \in S_2} Q(x)P(y|x) \leq 2 \sum_{(x,y) \in S_2} P(x)P(y|x) = 2P(S_2)
\]

Claim 2.3.13. \( Q(S_2) \leq 4\tilde{P}(S_2) + 16\alpha \)

Proof. We will show that \( D(\tilde{P}||Q) \leq 2\epsilon \) and then use Lemma (2.2.1) to conclude the claim. By definition,

\[
D(\tilde{P}||Q) = \sum_{(x,y) \in X \times Y} \tilde{P}(x, y) \log \left( \frac{\tilde{P}(x, y)}{Q(x, y)} \right)
\]

\[
= \sum_{(x,y), x \in X_2} Q(x)P(y|x) \log \left( \frac{Q(x)P(y|x)}{Q(x)Q(y|x)} \right) + \sum_{(x,y), x \notin X_2} \tilde{P}(x, y) \log \left( \frac{Q(x, y)}{Q(x, y)} \right)
\]

\[
= \sum_{(x,y), x \in X_2} \frac{Q(x)}{P(x)} \cdot P(x)P(y|x) \log \left( \frac{Q(x)P(y|x)}{Q(x)Q(y|x)} \right)
\]

Since for every \( x \in X_2, Q(x)/P(x) \leq 2 \) we obtain:

\[
D(\tilde{P}||Q) \leq 2 \sum_{(x,y), x \in X_2} P(x, y) \log \left( \frac{P(y|x)}{Q(y|x)} \right) \leq 2\mathbb{E}_{X \sim P_X} D(P_{Y|X}||Q_{Y|X}) \leq 4\epsilon
\]

where the last inequality follows from the assumption on \( P \). Applying Corollary (2.2.2) we obtain that \( Q(S_2) \leq 4\tilde{P}(S_2) + 16\alpha \)

\( \Box \)
Combining Claim (2.3.12) and Claim (2.3.13) we obtain

\[ Q(S_2) \leq 8P(S_2) + 16\alpha \leq 24\alpha \]

We now prove Lemma (2.3.1). Recall that,

\[ Q(S) = Q(S_1) + Q(S_2) + Q(S_3) \]

Combining Claim (2.3.9), Claim (2.3.11) and Claim (2.3.10) we obtain:

\[ Q(S) \leq 400(\alpha/\lambda)\log(17\alpha'/\alpha) + 40\alpha + 200(\alpha/\lambda)\log(17\alpha'/\alpha) \]
\[ \leq 600(\alpha/\lambda)\log(17\alpha'/\alpha) + 40\alpha \]
\[ \leq 640(\alpha/\lambda)\log(17\alpha'/\alpha) \]

That is

\[ \alpha'/\alpha \leq (640/\lambda)\log(17\alpha'/\alpha) \]

Recall that we assumed \( \alpha'/\alpha \geq 10^6 \cdot \frac{1}{\lambda} \cdot \log \frac{2}{\lambda} \). If

\[ \alpha'/\alpha = 10^6 \cdot \frac{1}{\lambda} \cdot \log \frac{2}{\lambda} \]

we obtain a contradiction since

\[ \alpha'/\alpha - (640/\lambda)\log(17\alpha'/\alpha) = \]
\[ 10^6 \cdot \frac{1}{\lambda} \cdot \log \frac{2}{\lambda} - (640/\lambda)\log \left( 17 \cdot 10^6 \cdot \frac{1}{\lambda} \cdot \log \frac{2}{\lambda} \right) = \]
\[ 10^6 \cdot \frac{1}{\lambda} \cdot \log \frac{2}{\lambda} - (640/\lambda)\log \left( 8.5 \cdot 10^6 \right) - (640/\lambda)\log \left( \frac{2}{\lambda} \right) - (640/\lambda)\log \left( \log \frac{2}{\lambda} \right) > 0 \]

To obtain a contradiction for all \( \alpha'/\alpha > 10^6 \cdot \frac{1}{\lambda} \cdot \log \frac{2}{\lambda} \) we now show that the derivative of the function

\[ f(z) = z - (640/\lambda)\log(17z) \]

is positive for \( z > 10^6 \cdot \frac{1}{\lambda} \cdot \log \frac{2}{\lambda} \) and hence, \( f(z) > 0 \) for all

\[ z \geq 10^6 \cdot \frac{1}{\lambda} \cdot \log \frac{2}{\lambda} \].
Substituting $z = \alpha' / \alpha$ we obtain a contradiction.

$$\frac{df}{dz} = 1 - \frac{640}{\lambda \ln 2} \cdot \frac{1}{z} > 0$$

where the last inequality follows since $z > 640/(\lambda \ln 2)$
Chapter 3

Parallel Repetition Theorems for Games on Expanders

3.1 Introduction

In this chapter, we study the case where the underlying distribution $P$, according to which the questions for the two provers are generated, is uniform over the edges of a (biregular) bipartite expander graph with sets of vertices $X, Y$. Let $M$ be the (normalized) adjacency matrix of the graph and denote by $1 - \lambda$ the second largest singular value of $M$. That is, $\lambda$ is the (normalized) spectral gap of the graph. (Note that the second largest singular value of $M$ is equal to the square root of the second largest eigenvalue of $MM^T$ and can be used to measure expansion in the same way).

We show that if $\lambda$ is the (normalized) spectral gap of the graph, the value of the repeated game is at most

$$(1 - \epsilon^2)^{\Omega(c(\lambda)n/s)},$$

where $c(\lambda) = \text{poly}(\lambda)$ (for general games); and at most

$$(1 - \epsilon)^{\Omega(c(\lambda)n)},$$
where \( c(\lambda) = \text{poly}(\lambda) \) (for projection games). In particular, for projection games we obtain a strong parallel repetition theorem (when \( \lambda \) is constant).

This gives a strong parallel repetition theorem for a large class of two prover games.

### 3.1.1 Related Works

**Strong Parallel Repetition Theorem for Free Projection Games:**

For games where the distribution \( P \) on \( X \times Y \) is a product distribution, we obtained bounds of \((1 - \epsilon^2)^{\Omega(n/s)}\) (for general games) and \((1 - \epsilon)^{\Omega(n)}\) (for projection games). We prove those bounds in Chapter 4. Note that product distributions can be viewed as distributions with maximal expansion.

**Almost Strong Parallel Repetition Theorem for Unique Games on Expanders:**

For the special case of unique games played on expander graphs, Arora, Khot, Kolla, Steurer, Tulsiani and Vishnoi previously proved an “almost” strong parallel repetition theorem (strong up to a polylogarithmic factor) \([6]\) (using \([20]\)). More precisely, they obtained a bound of \((1 - \epsilon^{2^{\log(1/\epsilon)}})^{\Omega(\lambda \cdot n)}\).

Safra and Schwartz \([36]\) obtained a strong parallel repetition theorem for the restricted class of games that satisfy all of the following: 1) the game is unique; 2) the roles of the two players are symmetric (that is, the game is played on a standard graph rather than a bipartite graph (note that the case of bipartite graph is more general)); and 3) the game is played on an expander graph that contains self loops with probability \(1/2\) on every vertex of the graph (that is, with probability \(1/2\) both players get the same question and are required to respond by the same answer). For such games, Safra and Schwartz obtained a bound of \((1 - \epsilon)^{\Omega(c(\lambda) \cdot n)}\), where \(c(\lambda) = \lambda / \log(2/\lambda)\).

Note that all these results are for the special case of unique games (played on expander graphs). In our work, we study the more general case of projection games.
Chapter 3. Parallel Repetition Theorems for Games on Expanders

Unique Games on Expander Graphs are Easy:

Several researchers studied the problem of strong parallel repetition with the motivation to use it for proving Khot’s unique games conjecture [26]. Recent results suggest that in order to prove the unique games conjecture, constructions that are exponential in $1/\epsilon$, such as constructions obtained using parallel repetition, are necessary [5, 4]. It turned out that a strong parallel repetition theorem (or even ‘close’ to strong) could be extremely helpful in studying the unique games conjecture. Since, a strong parallel repetition theorem is not true in general [33], it is interesting to try to prove it for general subclasses of games.

We note however that unique games on expander graphs are easy [6, 29, 28]. We hence don’t view the possible applications to the unique games conjecture as a major motivation for our result, since such application may require a substantial improvement of our results.

3.1.2 Techniques

Our proof goes along the general lines of the original proof of the parallel repetition theorem [32]. Unlike most recent works that went according to these lines, we are not able to use Holenstein’s new approach, as it results in a quadratic loss in $\epsilon$, that we cannot afford.

In previous results [32, 24, 31], a bound on the distance between a distribution “generated by the provers’ strategies” and the original distribution was derived using the relative entropy between the two distributions. This bound was then used to obtain a bound on the $\ell_1$ distance between those distributions. This was done using the fact that $\|P - Q\|_1 \leq O(\sqrt{D(P\|Q)})$ where $D(P\|Q)$ is the relative entropy between $P$ and $Q$. Since the bound is quadratic, there is a loss when using the $\ell_1$ norm instead of using directly the relative entropy. A recent counterexample shows that in the general case,
this loss is necessary [33].

We show that for the special case of games played on expander graphs, one can do the entire proof using relative entropy, without switching to $\ell_1$ norm. Our main technical contribution is a new lemma that may be interesting in its own right and can be stated roughly as follows. Let $Q$ be a uniform distribution on the edges of a bipartite expander graph with sets of vertices $X, Y$, and let $P$ be any distribution on the edges of that graph. If the conditional relative entropy $D(P_{X|Y}||Q_{X|Y})$ and the conditional relative entropy $D(P_{Y|X}||Q_{Y|X})$ are both at most $\epsilon$ then for any set $S \subset X \times Y$ with $P(S) \geq \epsilon$, we have that $Q(S) = O(P(S))$. This shows that if the two conditional relative entropies are small then the two distributions are “close” to each other, or, intuitively, if the two distributions are “close” from the point of view of an average vertex, then they are “close” to each other on the entire space. The proof of the lemma is not simple and uses several new ideas. We note, however, that a weaker lemma that allows a quadratic loss (that we cannot afford), would be much simpler.

Using the new lemma, the proof for the general case goes along the lines of the proof in [32]. The proof for the case of projection games requires the use of Rao’s ideas [31] together with additional new techniques. In particular, for the case of projection games we use once again the expansion properties of the underlying graph.

### 3.2 Preliminaries

#### 3.2.1 Notations

Recall the notations in Chapter 2 Section 2.1.1 and Section 2.3.1.

**Random Variables and Sets**

By slightly abusing notations, we will use capital letters to denote both sets and random variables distributed over these sets, and we will use lower case letters to denote values.
For example, $X, Y$ will denote sets as well as random variables distributed over these sets, and $x, y$ will denote values in these sets that the random variables can take. Nevertheless, it will always be clear from the context whether we are referring to sets or random variables. For a random variable $Z$ it will be convenient in some lemmas to think of $\Pr(Z)$ as a random variable.

**Random Variables and their Distributions**

For a random variable $X$, we denote by $P_X$ the distribution of $X$. For an event $U$ we use the notation $P_{X|U}$ to denote the distribution of $X|U$, that is, the distribution of $X$ conditioned on the event $U$. If $Z$ is an additional random variable that is fixed (e.g., inside an expression where an expectation over $Z$ is taken), we denote by $P_{X|Z}$ the distribution of $X$ conditioned on $Z$. In the same way, for two (or more) random variables $X, Y$, we denote their joint distribution by $P_{XY}$, and we use the same notations as above to denote conditional distributions. For example, for an event $U$, we write $P_{XY|U}$ to denote the distribution of $X, Y$ conditioned on the event $U$, i.e., $P_{XY|U}(x, y) = \Pr(X = x, Y = y|U)$. For two (or more) random variables $X, Y$ with distribution $P_{XY}$, we use the notation $P_X$ to denote the marginal distribution of $X$.

**The Game $G$**

We denote a game by $G$ and define $X$ to be the set of questions to prover 1, $Y$ to be the set of questions to prover 2 and $P_{XY}$ to be the joint distribution according to which the verifier chooses a pair of questions to the provers. We denote by $A$ the set of answers of prover 1 and by $B$ the set of answers of prover 2. We denote the acceptance predicate by $V$. A game $G$ with acceptance predicate $V$ and questions distribution $P_{XY}$ is denoted by $G(P_{XY}, V)$. As mentioned above, we also denote by $X, Y, A, B$ random variables distributed over $X, Y, A, B$ respectively. $X, Y$ will be the questions addressed to the two provers, distributed over the question sets $X$ and $Y$ respectively. Fixing a
strategy $f_a, f_b$ for the game $G$, we can also think of the answers $A$ and $B$ as random variables distributed over the answer sets $A$ and $B$ respectively.

The Game $G$ Repeated $n$ Times

For the game $G$ repeated $n$ times in parallel, $G^\otimes n = G(P_{X^n, Y^n}, V^\otimes n)$, the random variable $X_i$ denotes the question to prover 1 in coordinate $i$, and similarly, the random variable $Y_i$ denotes the question to prover 2 in coordinate $i$. We denote by $X^n$ the tuple $(X_1, \ldots, X_n)$ and by $Y^n$ the tuple $(Y_1, \ldots, Y_n)$. Fixing a strategy $f_a, f_b$ for $G^\otimes n$, the random variable $A_i$ denotes the answer of prover 1 in coordinate $i$, and similarly, the random variable $B_i$ denotes the answer of prover 2 in coordinate $i$. We denote by $A^n$ the tuple $(A_1, \ldots, A_n)$ and by $B^n$ the tuple $(B_1, \ldots, B_n)$. It will be convenient in some lemmas to denote $X^k = (X_{n-k+1}, \ldots, X_n)$, i.e., the last $k$ coordinates of $X^n$ and in the same way, $Y^k = (Y_{n-k+1}, \ldots, Y_n)$, $A^k = (A_{n-k+1}, \ldots, A_n)$ and $B^k = (B_{n-k+1}, \ldots, B_n)$. We also denote $X^{n-k} = (X_1, \ldots, X_{n-k})$, i.e., the first $n - k$ coordinates of $X^n$, and similarly, $Y^{n-k} = (Y_1, \ldots, Y_{n-k})$.

The Event $W_i$

For the game $G^\otimes n = G(P_{X^n, Y^n}, V^\otimes n)$ and a strategy $f_a : X^n \rightarrow A^n, f_b : Y^n \rightarrow B^n$ we can consider the joint distribution:

\[
P_{X^n, Y^n, A^n, B^n}(x^n, y^n, a^n, b^n) = \begin{cases} 
P_{X^n, Y^n}(x^n, y^n) & \text{if } a^n = f_a(x^n) \text{ and } b^n = f_b(y^n) \\ 0 & \text{otherwise} \end{cases}
\]

We define the event $W_i$ to be the event of winning the game in coordinate $i$, i.e., the event that the verifier accepts on coordinate $i$. Since the random variables $A^n$ and $B^n$ are functions of $X^n$ and $Y^n$ respectively, we can think of $W_i$ as an event in the random variables $X^n, Y^n$.

**Definition 3.2.1** (Projection Games). A Projection game is a game where for each pair
of questions $x, y$ there is a function $f_{xy}: B \to A$ such that $V(x, y, a, b)$ is satisfied if and only if $f_{xy}(b) = a$. 
3.3 Proof of Main Results

Recall the notations from Section (3.2.1). We define $W$ to be the event that the provers win all the games in the last $k$ coordinates. Recall that for a fixed $i$, $W_i$ is used for the event of winning the game in coordinate $i$. For a fixed $i \in [n-k]$ we define $M^{-i}$ in the following way. We let $D_1, \ldots, D_{n-k}$ be uniform and independent bits (either 0 or 1). For $1 \leq j \leq n-k$ we define

$$M_j = \begin{cases} 
X_j & \text{if } D_j = 0; \\
Y_j & \text{if } D_j = 1.
\end{cases}$$

Denote $M^{-i} := M_1, \ldots, M_{i-1}, M_{i+1}, \ldots, M_{n-k}$. For a fixed $D_1, \ldots, D_{n-k}$ the random variable $M_j$ is distributed over the set $X$ if $D_j = 0$ or over the set $Y$ if $D_j = 1$. Vaguely speaking, $M^{-i}$ is distributed over all possible ways to choose a question for exactly one of the provers in each coordinate $1 \leq j \leq n-k$ for $j \neq i$.

3.3.1 General Games

Recall that we consider a game $G(P_{XY}, V)$ played on a $(X, Y, d_X, d_Y, 1-\lambda)$-expander graph.

**Lemma 3.3.1.** For general games and the event $W$,

$$\mathbb{E}_{i \in [n-k]}(\mathbb{E}_{x^k, A^k, M^{-i}|W}\mathbb{E}_{x^k|X^k, A^k, M^{-i}, W}D(P_{Y_i|x_i,x^k, A^k, M^{-i}, W} \| P_{Y_i|x_i}) + \mathbb{E}_{x^k, A^k, M^{-i}|W}\mathbb{E}_{x^k|X^k, A^k, M^{-i}, W}D(P_{X_i|x_i,x^k, A^k, M^{-i}, W} \| P_{X_i|x_i})) \leq \frac{1}{n-k}(-\log(\Pr[W]) + k \log s)$$

where $s$ is the size of the answers set.

**Proof.** By [32] Claim (5.3) in the proof of Lemma (4.2) we obtain that for every $x^k \in$
3.3. PROOF OF MAIN RESULTS

\[X^k, y^k \in Y^k, a^k \in A^k,\]

\[
\mathbb{E}_{i \in [n-k]} \left( \mathbb{E}_{M^{-i}|x^k, y^k, a^k, W} \mathbb{E}_{X_i|x^k, y^k, a^k, M^{-i}, W} D(P|Y_i|x_i, x^k, y^k, a^k, M^{-i}, W || P|Y_i|x_i) + \mathbb{E}_{M^{-i}|x^k, y^k, a^k, W} \mathbb{E}_{Y_i|x^k, y^k, a^k, M^{-i}, W} D(P|X_i|Y_i, y^k, a^k, M^{-i}, W || P|X_i|Y_i) \right)
\leq \frac{1}{n-k} \left( -\log(\Pr[W|X^k = x^k, Y^k = y^k]) + \log(1/\Pr[A^k = a^k|X^k = x^k, Y^k = y^k, W]) \right)
\]

Using this, we obtain:

\[
\mathbb{E}_{X^k, Y^k, A^k|W} \mathbb{E}_{i \in [n-k]} \left( \mathbb{E}_{M^{-i}|x^k, y^k, A^k, W} \mathbb{E}_{X_i|x^k, y^k, A^k, M^{-i}, W} D(P|Y_i|x_i, x^k, y^k, A^k, M^{-i}, W || P|Y_i|x_i) + \mathbb{E}_{M^{-i}|x^k, y^k, A^k, W} \mathbb{E}_{Y_i|x^k, y^k, A^k, M^{-i}, W} D(P|X_i|Y_i, y^k, A^k, M^{-i}, W || P|X_i|Y_i) \right)
\leq \mathbb{E}_{X^k, Y^k, A^k|W} \cdot \frac{1}{n-k} \left( -\log(\Pr[W|X^k, Y^k]) + \log(1/\Pr[A^k|X^k, Y^k, W]) \right)
\leq \frac{1}{n-k} \cdot \log \left( \mathbb{E}_{X^k, Y^k|W} \cdot \frac{1}{\Pr[W|X^k, Y^k]} \right) + \frac{1}{n-k} \cdot \mathbb{E}_{X^k, Y^k|W} \log(1/\Pr[A^k|X^k, Y^k, W])
\leq -\frac{\log(\Pr[W])}{n-k} + \frac{1}{n-k} \cdot \mathbb{E}_{X^k, Y^k|W} \mathbb{E}_{A^k|X^k, Y^k, W} \log(1/\Pr[A^k|X^k, Y^k, W])
\leq -\frac{\log(\Pr[W])}{n-k} + \frac{1}{n-k} \cdot \mathbb{E}_{X^k, Y^k|W} H(P|A^k|X^k, Y^k, W)
\]

We use the trivial bound on the size of the support, namely, for every \(x^k, y^k\) we can bound \(|\text{supp}(P_{A^k|X^k=x^k, Y^k=y^k, W})| \leq |\text{supp}(P_{A^k})| \leq s^k\) where \(s\) is the size of the answers set. Using Fact 2.1.2 we can conclude the lemma. \(\square\)

**Lemma 3.3.2** (Main Lemma For General Games). Let \(G\) be a game with value \(1 - \epsilon\). Let \(T\) be the set of the last \(k\) coordinates, \((T = \{n - k + 1, \ldots, n\})\), let \(W\) be the event of the provers winning the games in those \(k\) coordinates. If \(\Pr(W) \geq 2^{-\epsilon(n-k)/4+\epsilon k \log s}\) where \(s\) is the size of the answers set and \(\epsilon' = 10^{-6}\epsilon\lambda/\log(2/\lambda)\), then there is \(i \notin T\) for which

\[
\Pr(W_i|W) \leq 1 - (1/2) \cdot 10^{-6} \cdot \epsilon \cdot \lambda / \log \frac{2}{\lambda} = 1 - \epsilon'/2
\]

**Proof.** For every \(i \in [n-k]\) and \(x^k, y^k, a^k, m^{-i}\), we will use a strategy for the game \(G(P_{X^n, Y^n, V^{\otimes n}})\) to obtain a strategy for the game \(G(P_{X_i, Y_i, X^k=x^k, Y^k=y^k, A^k=a^k, M^{-i}=m^{-i}, W, V})\).
Fix any strategy, $f_a, f_b$, for the game $G(P_{X^i,Y^i}, V^\otimes n)$, and apply the following to obtain a strategy for $G(P_{X^i,Y^i}|X^k=x^k, Y^k=y^k, A^k=a^k, M^{-i}=m^{-i}, W, V)$:

**Algorithm 3.3.3.** Protocol for $G(P_{X^i,Y^i}|X^k=x^k, Y^k=y^k, A^k=a^k, M^{-i}=m^{-i}, W, V)$ for fixed $x^k, y^k, a^k, m^{-i}, i$

1. When the game starts, prover 1 receives a question $x$ and prover 2 receives a question $y$ according to $P_{X^i,Y^i}|X^k=x^k, Y^k=y^k, A^k=a^k, M^{-i}=m^{-i}, W$. Define $X_i = x, Y_i = y$ (the provers will play this game in coordinate $i$).

2. Prover 1 randomly chooses the remaining questions (the questions that are not fixed in $m^{-i}$) according to $P_{X^{n-k}|X^k=x^k, Y^k=y^k, A^k=a^k, M^{-i}=m^{-i}, W, X_i=x}$. Denote those questions by $x^m$.

and Prover 2 randomly chooses the remaining questions (the questions that are not fixed in $m^{-i}$) according to $P_{Y^{n-k}|X^k=x^k, Y^k=y^k, A^k=a^k, M^{-i}=m^{-i}, W, Y_i=y}$. Denote those questions by $y^m$.

3. Prover 1 answers $[f_a(x^n)]_i$ and prove 2 answers $[f_b(y^n)]_i$.

**Remark 3.3.4.** Notice that in step 2, since $W$ is determined by $X^k, Y^k, A^k, B^k$, the joint distribution of $x^m, y^m$ is $P_{X^m,Y^m|X^k=x^k, Y^k=y^k, A^k=a^k, X_i=x, Y_i=y, M^{-i}=m^{-i}, W}$ since conditioned on $X_i$ (or $Y_i$) this distribution is a product distribution. Hence, the joint distribution of $x^n, y^n$ is

$$P_{X^n, Y^n|X^k=x^k, Y^k=y^k, A^k=a^k, M^{-i}=m^{-i}}$$

**Remark 3.3.5.** Notice that since Remark 3.3.4 holds, the probability of winning the game $G(P_{X^i,Y^i}|X^k=x^k, Y^k=y^k, A^k=a^k, M^{-i}=m^{-i}, W, V)$

is

$$\Pr(W_i|X^k = x^k, Y^k = y^k, A^k = a^k, M^{-i} = m^{-i}, W).$$
Remark 3.3.6. Notice that this is a randomized algorithm. However, it is well known that since any randomized algorithm is a convex combination of deterministic algorithms, there is a deterministic algorithm that achieves the same value as the randomized algorithm. Namely, there is a deterministic protocol for which the probability of winning the game

\[ G(P_{X,Y}|X^k=x^k, Y^k=y^k, A^k=a^k, M^{-i}=m^{-i}, W, V) \]

is at least

\[ \Pr(W_i|X^k = x^k, Y^k = y^k, A^k = a^k, M^{-i} = m^{-i}, W). \]

Using this remark we will think of this algorithm as a deterministic algorithm.

By Lemma 3.3.1 for a fixed strategy \( f_a, f_b \) for \( G(P_{X^*Y^*}, V^\otimes n) \),

\[
\mathbb{E}_{i \in [n-k]} \left( \mathbb{E}_{X^k,Y^k,A^k,M^{-i}|W} \mathbb{E}_{X_i|X^k,Y^k,A^k,M^{-i},W} D(P_{Y_i|X_i,X^k,Y^k,A^k,M^{-i},W} \| P_{Y_i|X_i}) + \mathbb{E}_{X^k,Y^k,A^k,M^{-i}|W} \mathbb{E}_{Y_i|X^k,Y^k,A^k,M^{-i},W} D(P_{X_i|Y_i,X^k,Y^k,A^k,M^{-i},W} \| P_{X_i|Y_i}) \right) \\
\leq \frac{1}{n-k} (- \log(\Pr[W]) + k \log s)
\]

By the assumption in the lemma, \( \Pr(W) \geq 2^{-\epsilon'(n-k)/4+k \log s} \). Therefore, it follows that:

\[
\mathbb{E}_{i \in [n-k]} \left( \mathbb{E}_{X^k,Y^k,A^k,M^{-i}|W} \mathbb{E}_{X_i|X^k,Y^k,A^k,M^{-i},W} D(P_{Y_i|X_i,X^k,Y^k,A^k,M^{-i},W} \| P_{Y_i|X_i}) + \mathbb{E}_{X^k,Y^k,A^k,M^{-i}|W} \mathbb{E}_{Y_i|X^k,Y^k,A^k,M^{-i},W} D(P_{X_i|Y_i,X^k,Y^k,A^k,M^{-i},W} \| P_{X_i|Y_i}) \right) \leq \epsilon'/4
\]

Assume by way of contradiction that for all \( i \in [n-k] \)

\[ \Pr(W_i|W) > 1 - \epsilon'/2. \]

Notice that since

\[ \Pr(W_i|W) = \mathbb{E}_{X^k,Y^k,A^k,M^{-i}|W} \Pr(W_i|X^k, Y^k, A^k, M^{-i}, W), \]

an equivalent assumption is that for all \( i \in [n-k] \),

\[ \mathbb{E}_{X^k,Y^k,A^k,M^{-i}|W} \Pr(\neg W_i|X^k, Y^k, A^k, M^{-i}, W) < \epsilon'/2. \]
Chapter 3. Parallel Repetition Theorems for Games on Expanders

By a simple averaging argument, there are \(i \in [n-k]\) and \(x^k, y^k, a^k, m^{-i}\) for which both equations hold:

\[
\mathbb{E}_{X_i|X^k=x^k, Y^k=y^k, A^k=a^k, M^{-i}=m^{-i}} W D(P_{Y_i|X_i, X^k=x^k, Y^k=y^k, A^k=a^k, M^{-i}=m^{-i}, W} || P_{Y_i|X_i}) +
\]

\[
\mathbb{E}_{Y_i|X^k=x^k, Y^k=y^k, A^k=a^k, M^{-i}=m^{-i}} W D(P_{X_i|Y_i, X^k=x^k, Y^k=y^k, A^k=a^k, M^{-i}=m^{-i}, W} || P_{X_i|Y_i}) \leq \epsilon'/2
\]

(3.3)

\[
\Pr(\neg W_i|X^k=x^k, Y^k=y^k, A^k=a^k, M^{-i}=m^{-i}, W) < \epsilon'
\]

(3.4)

For the strategy \(f_a, f_b\) and for \(x^k, y^k, a^k, i, m^{-i}\) for which both Equation (3.3) and Equation (3.4) hold consider the protocol suggested in Algorithm 3.3.3. Recall that by Remark 3.3.6 there is a deterministic protocol for which the provers win on coordinate \(i\) with probability at least

\[
\Pr(W_i|X^k=x^k, Y^k=y^k, A^k=a^k, M^{-i}=m^{-i}, W).
\]

Denote this deterministic protocol by \(h_a, h_b\). For \(h_a, h_b\), denote by \(R\) the set of all questions on which the provers err when playing according to this protocol. By the assumption in Equation (3.4)

\[
P_{X_i, Y_i|X^k=x^k, Y^k=y^k, A^k=a^k, M^{-i}=m^{-i}, W}(R) < \epsilon'.
\]

(3.5)

Combining Equation (3.5) with Equation (3.3), we can apply Lemma 2.3.1 to obtain \(P_{X_i, Y_i}(R) < \epsilon\). The provers can play \(h_a, h_b\) as a strategy for \(G(P_{X_i, Y_i}, V)\) and err only on questions in \(R\). Since \(P_{X_i, Y_i}(R) < \epsilon\), the value of \(G(P_{X_i, Y_i}, V) > 1 - \epsilon\) and since \(P_{X_i, Y_i} = P_{XY}\), the value of \(G(P_{XY}, V) > 1 - \epsilon\) which is a contradiction.

Recall that we consider a game \(G(P_{XY}, V)\) played on a \((X, Y, d_X, d_Y, 1 - \lambda)\)-expander graph.

**Theorem 6** (Parallel Repetition For General Games). For every game \(G\) with value \(1 - \epsilon\) where \(\epsilon < 1/2\), the value of \(G \otimes n\) is at most \((1 - \epsilon^2 \cdot c(\lambda))^{n/\log s}\) where \(s\) is the size of the answers set and \(c(\lambda) = (1/32)10^{-12} \lambda^2/((\log(\frac{2}{\lambda})))^2\).
3.3. PROOF OF MAIN RESULTS

Proof. (Of Theorem 6): We first show by induction that for every $k \leq n\epsilon'/(16 \log s)$ there is a set $T$ of $k$ coordinates for which,

$$\Pr(W) \leq (1 - \epsilon'/2)^k$$

where $W$ is the event of winning on all the coordinates in $T$ and $\epsilon' = 10^{-6} \epsilon \lambda/\log(2/\lambda)$.

Without loss of generality assume that $T$ is the set of the last $k$ coordinates, that is, $T = \{n - k + 1, \ldots, n\}$. For $k = 0$ the statement trivially holds. Assume by induction that for the set $T = \{n - k + 1, \ldots, n\}$, $\Pr(W) \leq (1 - \epsilon'/2)^k$. If $\Pr(W) \leq (1 - \epsilon'/2)^{k+1}$ then we are done. Otherwise

$$\Pr(W) > (1 - \epsilon'/2)^{k+1} \geq 2^{-(k+1)\epsilon'}$$

where we used the inequality $(1 - x) \geq 2^{-2x}$ for $0 \leq x \leq 1/2$. In order to use Lemma 3.3.2 we need to make sure that $\Pr(W) \geq 2^{-\epsilon'(n-k)/4+k\log s}$. It is enough to show that

$$2^{-(k+1)\epsilon'} \geq 2^{-\epsilon'(n-k)/4+k\log s}$$

or alternatively,

$$(k + 1)\epsilon' \leq \epsilon'(n - k)/4 - k \log s$$

After rearranging we obtain

$$k \leq \frac{\epsilon' n/4 - \epsilon'}{\log s + 1.25 \cdot \epsilon'}$$

For $|S| \geq 2$ and $n \geq 16$ it is enough that\footnote{We may assume that $|S| \geq 2$ since for $|S| = 1$ there is no dependency between the coordinates, therefore, there is a perfect parallel repetition and the theorem holds. We may assume that $n \geq 16$, otherwise, the theorem trivially holds.}

$$k \leq \frac{\epsilon' n}{16 \log s}.$$ 

Thus, for $k \leq n\epsilon'/(16 \log s)$ we can apply Lemma 3.3.2 to obtain that there is $i \notin T$ for which $\Pr(W_i|W) \leq 1 - \epsilon'/2$ therefore,

$$\Pr(W_i \land W) = \Pr(W) \cdot \Pr(W_i|W) \leq (1 - \epsilon'/2)^{k+1}$$
To complete the proof, set $k = n\epsilon'/(16 \log s)$ then as we showed, there is a set $T \subseteq [n], |T| = k$ for which:

$$\Pr(W_1 \land \ldots \land W_n) \leq \Pr(\bigwedge_{i \in T} W_i) \leq (1 - \epsilon'/2)^{n\epsilon'/(16 \log s)}$$

$$\leq (1 - (\epsilon')^2/32)^{n/\log s}$$

where the last inequality follows by the use of the inequality $(1 - x)^y \leq 1 - xy$ for every $0 \leq y \leq 1$ and $x \leq 1$.
3.3. PROOF OF MAIN RESULTS

3.3.2 Projection Games

Recall that we consider a game $G(P_{XY}, V)$ played on a $(X, Y, d_X, d_Y, 1 - \lambda)$-expander graph.

Lemma 3.3.7 (Main Lemma For Projection Games). Let $G$ be a projection game with value $1 - \epsilon$. Let $T$ be the set of the last $k$ coordinates, $(T = \{n - k + 1, \ldots, n\})$, let $W$ be the event of the provers winning the games in those $k$ coordinates. If $\Pr(W) \geq 1 - 10^{-12}\lambda$ and $(n - k) \geq (10^{-3} \cdot \epsilon' \cdot \lambda)^{-1}$ where $\epsilon' = 10^{-6}\epsilon\lambda/\log(2/\lambda)$, then there is $i \notin T$ for which

$$\Pr(W_i|W) \leq 1 - 10^{-8}\epsilon'$$

Proof. (Of Lemma (3.3.7)): If

$$\mathbb{E}_{i \notin T} \Pr(W_i|W) \leq 1 - 10^{-8}\epsilon'$$

then there is $i \notin T$ for which

$$\Pr(W_i|W) \leq 1 - 10^{-8}\epsilon'$$

and we are done. Thus we now assume that

$$\mathbb{E}_{i \notin T} \Pr(W_i|W) > 1 - 10^{-8}\epsilon'.$$

By the assumption in the lemma

$$\Pr(W) \geq 1 - 10^{-12}\lambda.$$

Claim 3.3.8. Given the event $W$, with probability of at least 80% over $x^k, y^k$ (that are chosen according to the conditional distribution $P_{X^k,Y^k|W}$) both equations hold:

$$\mathbb{E}_{i \notin T} \Pr(W_i|W, X^k = x^k, Y^k = y^k) > 1 - 10^{-7}\epsilon'$$ (3.6)

$$\Pr(W|X^k = x^k, Y^k = y^k) \geq 1 - 10^{-11}\lambda$$ (3.7)
Proof. Since
\[ \Pr(W) = \mathbb{E}_{X^k, Y^k} \Pr(W|X^k, Y^k) \geq 1 - 10^{-12}\lambda, \]
by a simple averaging argument, with probability of at least 90% over the \( x^k \in X^k, y^k \in Y^k \) that are chosen according to the distribution \( P_{X^kY^k} \),
\[ \Pr(W|X^k = x^k, Y^k = y^k) \geq 1 - 10^{-11}\lambda. \]
Let
\[ E_1 = \{ (x^k, y^k) | \Pr(W|X^k = x^k, Y^k = y^k) < 1 - 10^{-11}\lambda \}. \]
Thus, \( \Pr((X^k, Y^k) \in E_1) < 0.1 \) and
\[ \sum_{(x^k, y^k) \in E_1} \Pr(X^k = x^k, Y^k = y^k) \Pr(W|X^k = x^k, Y^k = y^k) < (1 - 10^{-11}\lambda) \cdot \sum_{(x^k, y^k) \in E_1} \Pr(X^k = x^k, Y^k = y^k) < 0.1 \cdot \Pr(W). \]
Hence,
\[ \sum_{(x^k, y^k) \in E_1} \Pr(X^k = x^k, Y^k = y^k|W) < 0.1 \tag{3.8} \]
Notice that
\[ \mathbb{E}_{i \in T} \Pr(W_i|W) = \mathbb{E}_{i \in T} \Pr(W_i \land W) / \Pr(W) \]
\[ = \mathbb{E}_{i \in T} \sum_{x^k \in X^k, y^k \in Y^k} \frac{\Pr(X^k = x^k, Y^k = y^k) \Pr(W|X^k = x^k, Y^k = y^k)}{\Pr(W)} \Pr(W_i|X^k = x^k, Y^k = y^k, W) \]
\[ = \mathbb{E}_{i \in T} \sum_{x^k \in X^k, y^k \in Y^k} \Pr(X^k = x^k, Y^k = y^k|W) \Pr(W_i|X^k = x^k, Y^k = y^k, W) \]
\[ = \sum_{x^k \in X^k, y^k \in Y^k} \Pr(X^k = x^k, Y^k = y^k|W) \mathbb{E}_{i \in T} \Pr(W_i|X^k = x^k, Y^k = y^k, W) \]
Denote
\[ E_2 = \{ (x^k, y^k) | \mathbb{E}_{i \in T} \Pr(W_i|W, X^k = x^k, Y^k = y^k) < 1 - 10^{-7}\epsilon' \}. \]
Thus,

\[ 1 - 10^{-8} \epsilon' < \mathbb{E}_{t \in T} \Pr(W_t | W) \]

\[ < (1 - 10^{-7} \epsilon') \cdot \sum_{(x^k, y^k) \in E_2} \Pr(X^k = x^k, Y^k = y^k | W) \]

\[ + \sum_{(x^k, y^k) \notin E_2} \Pr(X^k = x^k, Y^k = y^k | W)) \]

\[ = 1 - 10^{-7} \epsilon' \cdot \sum_{(x^k, y^k) \in E_2} \Pr(X^k = x^k, Y^k = y^k | W) \]

Thus we obtain that

\[ \sum_{(x^k, y^k) \in E_2} \Pr(X^k = x^k, Y^k = y^k | W) < 0.1. \]  \hspace{1cm} (3.9)

By Equation (3.8) and Equation (3.9) we obtain that with probability of at least 80% over all \( x^k \in X^k, y^k \in Y^k \) that are taken according to the distribution \( P_{X^kY^k|W} \), both Equation (3.6) and Equation (3.7) hold. \( \square \)

We now show that for projection games on expanders, if the probability of winning is high enough, then there is one answer that is obtained with high probability.

**Claim 3.3.9.** For every \( x^k \in X^k, y^k \in Y^k \) for which

\[ \Pr(W | X^k = x^k, Y^k = y^k) \geq 1 - 10^{-11} \lambda \]

there exists \( a^k \in A^k \) for which

\[ \Pr(A^k = a^k | X^k = x^k, Y^k = y^k) \geq 1/2 \]

**Proof.** Otherwise we can partition the set of answers \( A^k \) into two sets \( A', A'' \) such that

\[ \Pr(A^k \in A' | X^k = x^k, Y^k = y^k) \geq 1/4 \]

\[ \Pr(A^k \in A'' | X^k = x^k, Y^k = y^k) \geq 1/4 \]
Chapter 3. Parallel Repetition Theorems for Games on Expanders

Partition $X^{n-k}$ into two sets $X', X''$ that correspond to the answers $A', A''$ (when $X^k = x^k, Y^k = y^k$). That is, $x^{n-k} \in X'$ if on the last $k$ coordinates $f_a(x^{n-k}x^k)$ is an answer in $A'$; more formally, if

$$[f_a(x^{n-k}x^k)]^k \in A'.$$

Similarly, $x^{n-k} \in X''$ if

$$[f_a(x^{n-k}x^k)]^k \in A''.$$

Note that the probability for both $X', X''$ conditioned on $X^k = x^k, Y^k = y^k$ is at least 1/4. Partition $Y^{n-k}$ into two sets $Y', Y''$ according to the last $k$ coordinates of the answer, where $Y'$ corresponds to answers that project to answers in $A'$ and $Y''$ corresponds to answers that project to $A''$. That is, $y^{n-k} \in Y'$ if

$$f_{x^k y^k}([f_b(y^{n-k}y^k)]^k) \in A'$$

and $y^{n-k} \in Y''$ if

$$f_{x^k y^k}([f_b(y^{n-k}y^k)]^k) \in A''.$$

(where for $b^k \in B^k$, $f_{x^k y^k}(b^k)$ is the answer $a^k \in A^k$ that $b^k$ projects to)

Since the game is a projection game, the protocol err on both $X' \times Y''$ and $X'' \times Y'$ (when $X^k = x^k, Y^k = y^k$). We will now show that since the game is played on an expander graph, there must be many edges in $X' \times Y''$ or $X'' \times Y'$. We will examine paths of length two. In Claim (3.3.10), we will show that there are ‘many’ length two paths from $X'$ to $X''$ and then derive that there must be many edges in $(X' \times Y'') \cup (X'' \times Y')$.

For a game $G(P_{XY}, V)$ played on a $(X, Y, d_X, d_Y, 1-\lambda)$-expander graph $G$, denote by $M$ the $|X| \times |Y|$-adjacency matrix of $G$. The adjacency matrix of the graph is

$$M^{\otimes(n-k)} = M \otimes \cdots \otimes M.$$ 

This is the $|X^{n-k}| \times |Y^{n-k}|$-adjacency matrix of the $(d_X^{n-k}, d_Y^{n-k})$-bipartite graph

$$G^{\otimes(n-k)} = (X^{n-k} \cup Y^{n-k}, E').$$

53
where \((x^{n-k}, y^{n-k}) \in E'\) if and only if for every \(i \in \{1, \ldots, n - k\}\), \(P_{XY}(x_i, y_i) > 0\). Note that since the second normalized singular value of \(G\) is \(1 - \lambda\), the second normalized singular value of \(G^{\otimes(n-k)}\) is also \(1 - \lambda\). In this section we will only focus on \(G^{\otimes(n-k)}\) thus, for simplicity, we will denote \(d_X\) the degree of each \(x^{n-k} \in X^{n-k}\) and denote \(d_Y\) the degree of each \(y^{n-k} \in Y^{n-k}\) and also set \(G = G^{\otimes(n-k)}\).

Claim 3.3.10. The number of length two paths from \(X'\) to \(X''\) is at least

\[
\frac{1}{2} |X'|d_X \cdot d_Y \cdot \lambda
\]

Proof. Denote by \(M\) the \(|X| \times |Y|\)-adjacency matrix of \(G\) with normalized second-largest singular value \(1 - \lambda\). Thus, the \(|X| \times |X|\)-adjacency matrix of \(G^2\) is \(MM^T\) and \(G^2\) is a \(d = dXdY\) regular graph with second normalized eigenvalue of \((1 - \lambda)^2\).

Definition 3.3.11 (Edge Expansion). The edge expansion \(h(G)\) of a graph \(G = (V, E)\) is defined as

\[
h(G) = \min_{S \subseteq V, 1 \leq |S| \leq |V|/2} \frac{\partial(S)}{|S|}
\]

Where \(\partial(S)\) stands for the cardinality of the set of edges with exactly one endpoint in \(S\), namely, the number of edges between \(S\) and \(V \setminus S\).

Fact 3.3.12. For a \(d\) regular expander graph \(G\) with second eigenvalue \(1 - \lambda\),

\[
h(G) \geq \frac{1}{2}(d - d(1 - \lambda)) = \frac{1}{2}d\lambda.
\]

Hence,

\[
h(G^2) \geq \frac{1}{2}(d - d(1 - \lambda)^2) \geq \frac{1}{2}d\lambda
\]

See proof in [3], [14], [1]

By the expansion property of \(G^2\) the number of edges between \(^2\) \(X'\) and \(X''\) is at least \(|X'|h(G^2)\). By Fact (3.3.12) we obtain that the number of edges in \(G^2\) between \(X'\) and

---

\(^2\)We assume without loss of generality that \(|X'| \leq 1/2|X^{n-k}|\) otherwise we do the same argument on \(X''\).
Chapter 3. Parallel Repetition Theorems for Games on Expanders

$X''$ is at least

$$\frac{1}{2} |X'|d\lambda = \frac{1}{2} |X'|d_X \cdot d_Y \lambda.$$

Therefore, either, at least $\frac{1}{4} |X'|d_X \cdot d_Y \lambda$ of the length two paths from $X'$ to $X''$ go through vertices in $Y'$ (Case 1) or at least $\frac{1}{4} |X'|d_X \cdot d_Y \lambda$ of the length two paths from $X'$ to $X''$ go through vertices in $Y''$ (Case 2).

For every $y \in Y$ denote the number of edges from $y$ to vertices in $X'$ by $r'_y$ and the number of edges from $y$ to vertices in $X''$ by $r''_y$. Notice that the number of edges in $X' \times Y''$ is exactly:

$$\sum_{y \in Y''} r'_y$$

and the number of edges in $X'' \times Y'$ is exactly:

$$\sum_{y \in Y'} r''_y$$

In Case 1, using those notations, the number of length two paths from $X'$ to $X''$ that go through vertices in $Y'$ is

$$\sum_{y \in Y'} r'_y \cdot r''_y$$

Since for every $y \in Y$, $r'_y \leq d_Y$ and by our assumption,

$$d_Y \sum_{y \in Y'} r''_y \geq \sum_{y \in Y'} r'_y \cdot r''_y \geq \frac{1}{4} |X'|d_X \cdot d_Y \lambda$$

Thus, the number of edges in $X'' \times Y'$,

$$\sum_{y \in Y'} r''_y \geq \frac{1}{4} |X'|d_X \lambda.$$

In Case 2, using those notations, the number of length two paths from $X'$ to $X''$ that go through vertices in $Y''$ is

$$\sum_{y \in Y''} r'_y \cdot r''_y$$
Since for every \( y \in Y, r''_y \leq d_Y \) and by our assumption,
\[
d_Y \sum_{y \in Y''} r'_y \geq \sum_{y \in Y''} r'_y \cdot r''_y \geq \frac{1}{4} |X'| d_X \cdot d_Y \lambda
\]
Thus, the number of edges in \( X' \times Y'' \),
\[
\sum_{y \in Y''} r'_y \geq \frac{1}{4} |X'| d_X \lambda.
\]
Thus, either in \( X' \times Y'' \) or in \( X'' \times Y' \), the number of edges is at least \((1/4)|X'|d_X \lambda\).
Recall that \(|X'|/|X'| \leq 1/4\), therefore, the probability of winning the game conditioned on \( X^k = x^k, Y^k = y^k \) is at most
\[
1 - \frac{1}{16} \lambda
\]
which is a contradiction to the assumption on \( \Pr(W|X^k = x^k, Y^k = y^k) \).

By Claim (3.3.8) and Claim (3.3.9), with probability of at least 80% over \( x^k, y^k \) that are taken according to the distribution \( P_{X^k Y^k|W} \) there exists \( a^k \) such that the following equations hold:
\[
\mathbb{E}_{i \notin T} \Pr(W_i|W, X^k = x^k, Y^k = y^k) > 1 - 10^{-7} \epsilon'
\]  
(3.10)
\[
\Pr(W|X^k = x^k, Y^k = y^k) \geq 1 - 10^{-11} \lambda
\]  
(3.11)
\[
\Pr(A^k = a^k|X^k = x^k, Y^k = y^k) \geq 1/2
\]  
(3.12)

**Claim 3.3.13.** For every \( x^k, y^k, a^k \) that satisfy Equation (3.11) and Equation (3.12)
\[
\Pr(W|X^k = x^k, Y^k = y^k, A^k = a^k) \geq 1 - 10^{-10} \lambda
\]

**Proof.** Denote the event that the provers loose on the last \( k \) coordinates by \( \neg W \). Since,
\[
\Pr(\neg W|X^k = x^k, Y^k = y^k)
= \sum_{a^k \in A^k} \Pr(A^k = a^k|X^k = x^k, Y^k = y^k) \Pr(\neg W|X^k = x^k, Y^k = y^k, A^k = a^k)
\geq \Pr(A^k = a^k|X^k = x^k, Y^k = y^k) \Pr(\neg W|X^k = x^k, Y^k = y^k, A^k = a^k)
\geq (1/2) \Pr(\neg W|X^k = x^k, Y^k = y^k, A^k = a^k)
\]
Chapter 3. Parallel Repetition Theorems for Games on Expanders

By combining Equation (3.11) with Equation (3.12) we obtain

\[
\Pr(\neg W|X^k = x^k, Y^k = y^k, A^k = a^k) \leq 2 \cdot \Pr(\neg W|X^k = x^k, Y^k = y^k) < 2 \cdot 10^{-11} \lambda < 10^{-10} \lambda
\]

\[\phantom{\Pr(\neg W|X^k = x^k, Y^k = y^k, A^k = a^k) < 2 \cdot 10^{-11} \lambda < 10^{-10} \lambda}\]

Claim 3.3.14. With probability of at least 70\% over \( x^k, y^k \) that are taken according to the distribution \( P_{X^kY^k|W} \), there exists \( a^k \) that satisfies,

\[
\mathbb{E}_{i \notin T} \Pr(W_i|W, X^k = x^k, Y^k = y^k, A^k = a^k) > 1 - 10^{-6} \epsilon'
\]

Proof. Denote the event that the provers loose in the game in coordinate \( i \) by \( \neg W_i \).

Notice that

\[
\Pr(\neg W_i|W, X^k = x^k, Y^k = y^k) = \sum_{a^k \in A^k} \Pr(\neg W_i \land (A^k = a^k)|W, X^k = x^k, Y^k = y^k)
\]

\[
= \sum_{a^k \in A^k} \Pr(A^k = a^k|W, X^k = x^k, Y^k = y^k) \cdot \Pr(\neg W_i|W, X^k = x^k, Y^k = y^k, A^k = a^k)
\]

\[
= \sum_{a^k \in A^k} \frac{\Pr(A^k = a^k|X^k = x^k, Y^k = y^k) \cdot \Pr(W|X^k = x^k, Y^k = y^k, A^k = a^k)}{\Pr(W|X^k = x^k, Y^k = y^k)} \cdot \Pr(\neg W_i|W, X^k = x^k, Y^k = y^k, A^k = a^k)
\]

\[\phantom{= \sum_{a^k \in A^k} \frac{\Pr(A^k = a^k|X^k = x^k, Y^k = y^k) \cdot \Pr(W|X^k = x^k, Y^k = y^k, A^k = a^k)}{\Pr(W|X^k = x^k, Y^k = y^k)}} \tag{3.13}\]

With probability of at least 80\% over \( x^k, y^k \) that are taken according to the distribution \( P_{X^kY^k|W} \) there exists \( a^k \) for which both Equation (3.12) and Claim (3.3.13) hold. Thus, using Equation (3.13), Claim (3.3.13) and Equation (3.12), for those \( x^k, y^k, a^k \) and for every \( i \notin T \):

\[
\Pr(\neg W_i|W, X^k = x^k, Y^k = y^k) \\
\geq (1/2) \cdot (1 - 10^{-10} \lambda) \cdot \Pr(\neg W_i|W, X^k = x^k, Y^k = y^k, A^k = a^k)
\]

\[
\geq (1/4) \cdot \Pr(\neg W_i|W, X^k = x^k, Y^k = y^k, A^k = a^k)
\]

Therefore,

\[
\mathbb{E}_{i \notin T} \Pr(\neg W_i|W, X^k = x^k, Y^k = y^k, A^k = a^k) \leq 4 \cdot 10^{-7} \epsilon' < 10^{-6} \epsilon'
\]

57
Hence the claim follows. \(\square\)

Hence, there are \(x^k, y^k, a^k\) for which all of the following equations hold:

\[
\mathbb{E}_{i \in T} \Pr(W_i | W, X^k = x^k, Y^k = y^k, A^k = a^k) > 1 - 10^{-6} \epsilon'
\]  
(3.14)

\[
\Pr(W | X^k = x^k, Y^k = y^k) \geq 1 - 10^{-11} \lambda
\]  
(3.15)

\[
\Pr(W | X^k = x^k, Y^k = y^k, A^k = a^k) \geq 1 - 10^{-10} \lambda
\]

\[
\Pr(A^k = a^k | X^k = x^k, Y^k = y^k) \geq 1/2
\]

By [32] Claim (5.3) in the proof of Lemma (4.2) we obtain that for every \(x^k \in X^k, y^k \in Y^k, a^k \in A^k\),

\[
\mathbb{E}_{i \in [n-k]} \left( \mathbb{E}_{M^{-i} | x^k, y^k, a^k, W} \mathbb{E}_{X_i, x_i, y_i, a_i, M^{-i}, W} D(P_{Y_i | X_i, x_i, y_i, a_i, M^{-i}, W} || P_{Y_i | X_i}) + \mathbb{E}_{M^{-i} | x^k, y^k, a^k, W} \mathbb{E}_{Y_i | x_i, y_i, a_i, M^{-i}, W} D(P_{X_i | Y_i, x_i, y_i, a_i, M^{-i}, W} || P_{X_i | Y_i}) \right) \\
\leq \frac{1}{n-k} \left( - \log(\Pr[W | X^k = x^k, Y^k = y^k]) + \log(1 / \Pr[A^k = a^k | X^k = x^k, Y^k = y^k, W]) \right)
\]  
(3.16)

Since,

\[
\Pr[A^k = a^k | X^k = x^k, Y^k = y^k, W]
\]

\[
= \frac{\Pr[A^k = a^k | X^k = x^k, Y^k = y^k] \cdot \Pr[W | X^k = x^k, Y^k = y^k, A^k = a^k]}{\Pr[W | X^k = x^k, Y^k = y^k]}
\]

\[
\geq \Pr[A^k = a^k | X^k = x^k, Y^k = y^k] \cdot \Pr[W | X^k = x^k, Y^k = y^k, A^k = a^k]
\]

\[
\geq 1/2 \cdot (1 - 10^{-10} \lambda)
\]

and by applying both Equation (3.15) and the assumption

\[
(n - k) \geq (10^{-3} \cdot \epsilon' \cdot \lambda)^{-1},
\]

we obtain:

Equation (3.16) \leq (10^{-3} \cdot \epsilon' \cdot \lambda) \left( - \log(1 - 10^{-11} \lambda) - \log(0.5 \cdot (1 - 10^{-10} \lambda)) \right) < 10^{-2} \epsilon'
where the last inequality follows by using that $-\log(1 - x) \leq 1$ for $x < 1/2$ (recall that $\lambda < 1$). Since

$$\Pr(W_i|W, X^k = x^k, Y^k = y^k, A^k = a^k) =$$

$$\mathbb{E}_{M^{-i}|W, X^k = x^k, Y^k = y^k, A^k = a^k} \Pr(W_i|W, X^k = x^k, Y^k = y^k, A^k = a^k, M^{-i}),$$

By Equation (3.14) we can fix $i \in \{n - k\}$ and $m^{-i} \in M^{-i}$ for which

$$\Pr(W_i|W, X^k = x^k, Y^k = y^k, A^k = a^k, M^{-i} = m^{-i}) > 1 - \epsilon' \quad (3.17)$$

and

$$\mathbb{E}_{X_i|X^k = x^k, Y^k = y^k, A^k = a^k, M^{-i} = m^{-i}, W} \mathcal{D}(P_{Y_i|X_i, X^k = x^k, Y^k = y^k, A^k = a^k, M^{-i} = m^{-i}, W} || P_{Y_i|X_i}) + (3.18)$$

$$\mathbb{E}_{Y_i|X^k = x^k, Y^k = y^k, A^k = a^k, M^{-i} = m^{-i}, W} \mathcal{D}(P_{X_i|Y_i, X^k = x^k, Y^k = y^k, A^k = a^k, M^{-i} = m^{-i}, W} || P_{X_i|Y_i}) < 10^{-1} \epsilon'$$

For the strategy $f_a, f_b$ and for $x^k, y^k, a^k, i, m^{-i}$ for which both Equation (3.17) and Equation (3.18) hold, consider the protocol suggested in Algorithm 3.3.3. Recall that by Remark 3.3.6 there is a deterministic protocol for which the provers win on coordinate $i$ with probability at least

$$\Pr(W_i|X^k = x^k, Y^k = y^k, A^k = a^k, M^{-i} = m^{-i}, W).$$

Denote this deterministic protocol by $h_a, h_b$. For $h_a, h_b$, denote by $R$ the set of all questions on which the provers err when playing according to this protocol. By the assumption in Equation (3.17)

$$P_{X_i, Y_i|X^k = x^k, Y^k = y^k, A^k = a^k, M^{-i} = m^{-i}, W}(R) < \epsilon'. \quad (3.19)$$

Combining Equation (3.19) with Equation (3.18), we can apply Lemma 2.3.1 to obtain $P_{X_i, Y_i}(R) < \epsilon$. The provers can play $h_a, h_b$ as a strategy for $G(P_{X_i, Y_i}, V)$ and err only on questions in $R$. Since $P_{X_i, Y_i}(R) < \epsilon$, the value of $G(P_{X_i, Y_i}, V) > 1 - \epsilon$ and since $P_{X_i, Y_i} = P_{XY}$, the value of $G(P_{XY}, V) > 1 - \epsilon$ which is a contradiction.
Recall that we consider a game $G(P_{XY}, V)$ played on a $(X, Y, d_X, d_Y, 1 - \lambda)$-expander graph.

**Theorem 7** (Parallel Repetition For Projection Games). *For every projection game $G$ with value $1 - \epsilon$ where $\epsilon < 1/2$, the value of $G^\otimes n$ is at most $(1 - \epsilon)^{\text{poly}(\lambda) \cdot n}$*

**Proof Sketch**  By inductive application of Lemma (3.3.7) as in the proof of Theorem (6) we can reduce the value of the game from $1 - \epsilon$ to $1 - \Omega(\lambda)$ by $O(1/(\epsilon \cdot \text{poly}(\lambda)))$ repetitions. Then by applying Rao’s bound [31] we can further reduce the value of the game to any constant by $O(1/\text{poly}(\lambda))$ repetitions of this protocol (that is, a total number of $O(1/(\epsilon \cdot \text{poly}(\lambda)))$ repetitions.

Note that in order to apply Lemma (3.3.7) we need $n$ to be large enough. Nevertheless, as suggested in [31], it can be shown that if the theorem was false for small $n$ it would not hold for big $n$. If there was a strategy with high success probability for small $n$ this strategy could be repeated in parallel to give a contradiction for large $n$.

**Proof.** (of Theorem (7)): By the observation above, we may assume throughout the proof that

$$n \geq 10^4(\epsilon')^{-1} \lambda^{-1}.$$  

We first show by induction that for every $k \leq \log(2/\lambda)\epsilon^{-1}$ there is a set $T \subseteq [n]$ of $k$ coordinates ($|T| = k$) for which

$$\Pr(W) \leq \left(1 - 10^{-14} \cdot \epsilon \cdot \lambda / \log 2 \lambda \right)^k$$

where $W$ is the event of winning on all the coordinates in $T$. For $k = 0$ the statement trivially holds. Assume by induction that there is a set $T$ of size $k$ for which

$$\Pr(W) \leq \left(1 - 10^{-14} \cdot \epsilon \cdot \lambda / \log 2 \lambda \right)^k.$$  

---

$^3$where $\epsilon' = 10^{-6} \cdot \epsilon \lambda / \log(2/\lambda)$
If
\[ \Pr(W) \leq \left( 1 - 10^{-14} \cdot \epsilon \cdot \lambda / \log \frac{2}{\lambda} \right)^{k+1} \]
then we are done. Otherwise
\[ \Pr(W) > \left( 1 - 10^{-14} \cdot \epsilon \cdot \lambda / \log \frac{2}{\lambda} \right)^{k+1} \geq 2^{-2(k+1)\cdot 10^{-14} \cdot \epsilon \cdot \lambda / \log \frac{2}{\lambda}} \]
where we used the inequality \((1 - x) \geq 2^{-2x} \) for \(0 \leq x \leq 1/2\). In order to use Lemma 3.3.7 we need to make sure that
\[ \Pr(W) \geq 1 - 10^{-12} \lambda \]
and that
\[ (n - k) \geq (10^{-3} \cdot \epsilon' \cdot \lambda)^{-1} \]
(and we assume without loss of generality that \(T = \{n - k + 1, \ldots, n\}\)).

Since \((1 - x) \leq 2^{-x} \) for \(0 \leq x \leq 1/2\), it is enough to show that
\[ 2^{-2(k+1)\cdot 10^{-14} \cdot \epsilon \cdot \lambda / \log \frac{2}{\lambda}} \geq 2^{-10^{-12} \lambda}. \]
or alternatively,
\[ 2(k + 1)\epsilon \cdot 10^{-2} / \log \frac{2}{\lambda} \leq 1 \]
For \(k \leq \log(2/\lambda)\epsilon^{-1}\) this inequality holds.

We showed that the value of the game \(G^{\otimes 10^4(\epsilon')^{-1}\lambda^{-1}}\) is at most
\[ \left( 1 - 10^{-14} \cdot \epsilon \cdot \lambda / \log \frac{2}{\lambda} \right)^{\log(2/\lambda)\epsilon^{-1}} \leq 1 - 10^{-15} \cdot \lambda \)
(3.20)

We now state Rao’s theorem ([31] Theorem 4):

There is a universal constant \(c > 0\) such that if \(G\) is a projection game with
value at most \(1 - \epsilon\), the value of \(G^{\otimes n}\) is at most \((1 - c\epsilon^2)^n\)

We think of the game \(G\) played \(n\) times in parallel as the game \(G^{\otimes 10^4(\epsilon')^{-1}\lambda^{-1}}\) played
\(n \cdot 10^{-4}(\epsilon')\lambda\) times in parallel. Thus combining Equation (3.20) with Rao’s theorem, we
obtain that the value of \(G^{\otimes n}\) is at most \((1 - \Omega(\lambda^2))^{n \cdot 10^{-4}(\epsilon')\lambda} \leq (1 - \epsilon \cdot \text{poly}(\lambda))^n\)
3.3. PROOF OF MAIN RESULTS
Chapter 4

Parallel Repetition Theorems for Free Games

4.1 Introduction

In this chapter we study the case where the distribution according to which the questions to the provers are generated is a product distribution. That is, if $P_{XY} = P_X \cdot P_Y$. We refer to a game that satisfy the condition above as a free game.

Many researchers studied the problem of whether there exists a strong parallel repetition theorem in the general case or at least in some important special cases. Namely, is it the case that for a given game $G$ of value $1 - \epsilon$, say, for $\epsilon < 1/2$, the value of $G^\otimes n$ is at most $(1 - \epsilon)^{\Omega(n/\log s)}$? This question was motivated by connections to hardness of approximation as well as connections to problems in geometry [19], [36]. A recent result of Raz [33] showed a counterexample for the general case, as well as for the case of projection games, unique games and XOR games. Raz [33] showed that there is an example of a XOR game (thus also projection game and unique game) of value $1 - \epsilon$ such that for large enough $n$, the value of the game is at least $(1 - \epsilon^2)^{O(n)}$. For some extensions, generalization and applications see Barak, Hardt, Haviv, Rao, Regev and Steurer [8],
4.1. **INTRODUCTION**

Kindler, O’Donnell, Rao and Wigderzon [27] and Alon and Klartag [2].

Other related results: For the special case of unique games played on expander graphs Arora, Khot, Kolla, Steurer, Tulsiani and Vishnoi [6] proved an “almost” strong parallel repetition theorem (strong up to a polylogarithmic factor). For the special case of games where the roles of the two players are symmetric and the game is played on an expander graph that contains a self loop on every vertex, Safra and Schwartz [36] showed that $O(1/\epsilon)$ repetitions are sufficient to reduce the value of the game from $1 - \epsilon$ to some constant.

In this chapter we prove a strong parallel repetition theorem for free projection games and we improve the known bound for every free game. More precisely:

1. For every **Free game** of value $\leq (1 - \epsilon)$ for $\epsilon < 1/2$, the value of $G^\otimes n$ is at most $(1 - \epsilon^2)^\Omega(n/\log s)$

2. For every **Free Projection game** of value $\leq (1 - \epsilon)$ for $\epsilon < 1/2$, the value of $G^\otimes n$ is at most $(1 - \epsilon)^\Omega(n)$

**Techniques**

The main technical contribution of this chapter is the ability to work throughout the whole proof with relative entropy without the need to switch to $\ell_1$ norm. In previous results [32], [24], [31] a bound on the distance between a distribution “generated by the provers’ strategies” and the original distribution was derived using the relative entropy between the two distributions. This bound was then used to obtain a bound on the $\ell_1$ distance between those distributions. This was done using the fact that $\|P - Q\|_1 \leq O(\sqrt{D(P\|Q)})$ where $D(P\|Q)$ is the relative entropy between $P$ and $Q$. Since the bound is quadratic, there is a loss when using the $\ell_1$ norm instead of using directly the relative entropy. We show that for the special case of free games one can redo the whole proof using relative entropy, without switching to $\ell_1$ norm. We bound the value of a game by
using our Corollary 2.2.2 (that might be useful for other applications). We note that since we are only considering free games, the proof is simpler than the one for general games and we do not use much of the machinery used in previous results, e.g., [32], [24], [31].

4.2 Preliminaries

4.2.1 Notations

Recall the notations in Chapter 2 Section 2.1.1.

Random Variables and Sets

By slightly abusing notations, we will use capital letters to denote both sets and random variables distributed over these sets, and we will use lower case letters to denote values. For example, $X, Y$ will denote sets as well as random variables distributed over these sets, and $x, y$ will denote values in these sets that the random variables can take. Nevertheless, it will always be clear from the context whether we are referring to sets or random variables. For a random variable $Z$ it will be convenient in some lemmas, such as Lemma 4.3.5, to think of $\Pr(Z)$ as a random variable.

Random Variables and their Distributions

For a random variable $X$, we denote by $P_X$ the distribution of $X$. For an event $U$ we use the notation $P_{X|U}$ to denote the distribution of $X|U$, that is, the distribution of $X$ conditioned on the event $U$. If $Z$ is an additional random variable that is fixed (e.g., inside an expression where an expectation over $Z$ is taken), we denote by $P_{X|Z}$ the distribution of $X$ conditioned on $Z$. In the same way, for two (or more) random variables $X, Y$, we denote their joint distribution by $P_{XY}$, and we use the same notations as above to denote conditional distributions. For example, for an event $U$, we write $P_{XY|U}$ to denote the
distribution of $X, Y$ conditioned on the event $U$, i.e., $P_{XY|U}(x, y) = \Pr(X = x, Y = y|U)$.

For two (or more) random variables $X, Y$ with distribution $P_{XY}$, we use the notation $P_X$ to denote the marginal distribution of $X$.

**The Game $G$**

We denote a game by $G$ and define $X$ to be the set of questions to prover 1, $Y$ to be the set of questions to prover 2 and $P_{XY}$ to be the joint distribution according to which the verifier chooses a pair of questions to the provers. We denote by $A$ the set of answers of prover 1 and by $B$ the set of answers of prover 2. We denote the acceptance predicate by $V$. A game $G$ with acceptance predicate $V$ and questions distribution $P_{XY}$ is denoted by $G(P_{XY}, V)$. As mentioned above, we also denote by $X, Y, A, B$ random variables distributed over $X, Y, A, B$ respectively. $X, Y$ will be the questions addressed to the two provers, distributed over the question sets $X$ and $Y$ respectively. Fixing a strategy $f_a, f_b$ for the game $G$, we can also think of the answers $A$ and $B$ as random variables distributed over the answer sets $A$ and $B$ respectively.

**The Game $G$ Repeated $n$ Times**

For the game $G$ repeated $n$ times in parallel, $G^\otimes n = G(P_{X^nY^n}, V^\otimes n)$, the random variable $X_i$ denotes the question to prover 1 in coordinate $i$, and similarly, the random variable $Y_i$ denotes the question to prover 2 in coordinate $i$. We denote by $X^n$ the tuple $(X_1, \ldots, X_n)$ and by $Y^n$ the tuple $(Y_1, \ldots, Y_n)$. Fixing a strategy $f_a, f_b$ for $G^\otimes n$, the random variable $A_i$ denotes the answer of prover 1 in coordinate $i$, and similarly, the random variable $B_i$ denotes the answer of prover 2 in coordinate $i$. We denote by $A^n$ the tuple $(A_1, \ldots, A_n)$ and by $B^n$ the tuple $(B_1, \ldots, B_n)$. It will be convenient in some lemmas to denote $X^k = (X_{n-k+1}, \ldots, X_n)$, i.e., the last $k$ coordinates of $X^n$ and in the same way, $Y^k = (Y_{n-k+1}, \ldots, Y_n)$, $A^k = (A_{n-k+1}, \ldots, A_n)$ and $B^k = (B_{n-k+1}, \ldots, B_n)$. We also denote $X^{n-k} = (X_1, \ldots, X_{n-k})$, i.e., the first $n - k$ coordinates of $X^n$, and similarly, $Y^{n-k} = (Y_1, \ldots, Y_{n-k})$. We also denote $X^{n-k} = (X_1, \ldots, X_{n-k})$, i.e., the first $n - k$ coordinates of $X^n$, and similarly, $Y^{n-k} = (Y_1, \ldots, Y_{n-k})$. We also denote $X^{n-k} = (X_1, \ldots, X_{n-k})$, i.e., the first $n - k$ coordinates of $X^n$, and similarly, $Y^{n-k} =$
(Y_1, \ldots, Y_{n-k}). For fixed $$i \in [n - k]$$, we denote $$X^m = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_{n-k})$$, i.e., $$X^{n-k}$$ without $$X_i$$, and similarly, $$Y^m = (Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_{n-k})$$.

The Event $$W_i$$

For the game $$G^\otimes n = G(P_{X^nY^n}, V^\otimes n)$$ and a strategy $$f_a : X^n \to A^n, f_b : Y^n \to B^n$$ we can consider the joint distribution:

$$P_{X^n,Y^n,A^n,B^n}(x^n, y^n, a^n, b^n) = \begin{cases} P_{X^n,Y^n}(x^n, y^n) & \text{if } a^n = f_a(x^n) \text{ and } b^n = f_b(y^n) \\ 0 & \text{otherwise} \end{cases}$$

We define the event $$W_i$$ to be the event of winning the game in coordinate $$i$$, i.e., the event that the verifier accepts on coordinate $$i$$. Since the random variables $$A^n$$ and $$B^n$$ are functions of $$X^n$$ and $$Y^n$$ respectively, we can think of $$W_i$$ as an event in the random variables $$X^n, Y^n$$.

### 4.2.2 Special Types of Games

**Definition 4.2.1** (Free Games). A game is Free if the distribution of the questions is a product distribution, i.e., $$P_{XY} = P_X \times P_Y$$

**Definition 4.2.2** (Projection Games). A Projection game is a game where for each pair of questions $$x, y$$ there is a function $$f_{xy} : B \to A$$ such that $$V(x, y, a, b)$$ is satisfied if and only if $$f_{xy}(b) = a$$.

### 4.3 Our Results

We prove the following theorems:

**Theorem 8** (Parallel Repetition For Free Games). For every game $$G$$ with value $$1 - \epsilon$$ where $$\epsilon < 1/2$$ and $$P_{XY} = P_X \times P_Y$$ (the questions are distributed according to some product distribution), the value of $$G^\otimes n$$ is at most $$(1 - \epsilon^2/9)^{n/(18 \log s + 3)}$$
4.3. OUR RESULTS

Theorem 9 (Strong Parallel Repetition For Free Projection Games). For every projection game $G$ with value $1 - \epsilon$ where $\epsilon < 1/2$ and $P_{XY} = P_X \times P_Y$ (the questions are distributed according to some product distribution), the value of $G^\otimes n$ is at most $(1 - \epsilon/9)^{(n/33) - 1}$.

4.3.1 Technical Lemma

Recall Lemma 2.2.1. For the convenience of the reader we state the lemma in this section. For the proof of the lemma See Chapter 2 Section 2.2.

Lemma 4.3.1. For every $0 \leq p, q \leq 1$ define binary distributions $P = (p, 1 - p)$ and $Q = (q, 1 - q)$, over $\{0, 1\}$, if $D(P \parallel Q) \leq \delta$ and $p < \delta$ then

$$q \leq 4\delta$$

Corollary 4.3.2. For every probability distributions $P, Q$ over the same sample space $\Omega$ and for every $T \subseteq \Omega$, if $D(P \parallel Q) \leq \delta$ and $P(T) \leq \delta$ then $Q(T) \leq 4\delta$
4.3.2 Main Lemmas

We now state the main lemmas for general product distribution games.

Recall that for a coordinate \(i\), \(W_i\) is the event of the provers winning the game played in this coordinate.

**Lemma 4.3.3** (Main Lemma For General Free Games). Let \(G\) be a free game with value \(1 - \epsilon\). For any set \(T\) of \(k\) coordinates, \((T \subseteq [n] \text{ and } |T| = k)\), let \(W\) be the event of the provers winning the games in those \(k\) coordinates. If \(\Pr(W) \geq 2^{-(n-k)/9 + k \log s}\) where \(s\) is the size of the answers set, then there is \(i \notin T\) for which

\[
\Pr(W_i|W) \leq 1 - \frac{\epsilon}{9}
\]

**Lemma 4.3.4** (Main Lemma For Free Projection Games). Let \(G\) be a free projection game with value \(1 - \epsilon\). For any set \(T\) of \(k\) coordinates, \((T \subseteq [n] \text{ and } |T| = k)\), let \(W\) be the event of the provers winning the games in those \(k\) coordinates. If \(\Pr(W) \geq 2^{-(n-k)/144}\) and \(n - k \geq (48/\epsilon) \log(8/\epsilon)\) then there is \(i \notin T\) for which

\[
\Pr(W_i|W) \leq 1 - \frac{\epsilon}{9}
\]

In the lemmas below we assume without loss of generality that the set \(T\) of \(k\) coordinates is the set of the last \(k\) coordinates. Recall that \(P_{X_nY^n} = P_{XY} \times \cdots \times P_{XY}\) \(n\)-times. Recall that \(X^k = (X_{n-k+1}, \ldots, X_n)\), i.e., the last \(k\) coordinates of \(X^n\) and in the same way, \(Y^k = (Y_{n-k+1}, \ldots, Y_n)\), \(A^k = (A_{n-k+1}, \ldots, A_n)\) and \(B^k = (B_{n-k+1}, \ldots, B_n)\). Recall that \(X^{n-k} = (X_1, \ldots, X_{n-k})\), i.e., the first \(n-k\) coordinates of \(X^n\), and similarly, \(Y^{n-k} = (Y_1, \ldots, Y_{n-k})\).

**Lemma 4.3.5.** For any event\(^1\) \(U\), the following holds:

\[
\mathbb{E}_{X^k,Y^k,A^k|U,D} \left( P_{X_{n-k}Y_{n-k}|X^k,Y^k,A^k,U} \| P_{X_{n-k}Y_{n-k}} \right) \leq \log \left( \frac{1}{\Pr(U)} \right) + \mathbb{E}_{X^k,Y^k|U} H(P_{A^k|X^k,Y^k,U})
\]

\(^1\)We will use the lemma for events that depend only on \(X^k, Y^k, A^k, B^k\), e.g., we will use it for the event \(W\), see definition in Lemma 4.3.3
4.3. OUR RESULTS

Proof. Since $P_{X^n,Y^n} = P_{XY} \times \cdots \times P_{XY}$ $n$-times,

\[
\mathbb{E}_{X^k,Y^k,A^k|U} D \left( P_{X^{n-k},Y^{n-k}|X^k,Y^k,A^k,U} \| P_{X^{n-k},Y^{n-k}} \right)
\]

\[
= \mathbb{E}_{X^k,Y^k,A^k|U} \left( P_{X^{n-k},Y^{n-k}|X^k,Y^k,A^k,U} \| P_{X^{n-k},Y^{n-k}|X^k,Y^k} \right)
\]

\[
= \mathbb{E}_{X^k,Y^k,A^k|U} \mathbb{E}_{X^{n-k},Y^{n-k}|X^k,Y^k,A^k,U} \log \left( \frac{P_{X^{n-k},Y^{n-k}|X^k,Y^k,A^k,U}}{P_{X^{n-k},Y^{n-k}|X^k,Y^k}} \right)
\]

\[
= \mathbb{E}_{X^k,Y^k,A^k|U} \mathbb{E}_{X^{n-k},Y^{n-k}|X^k,Y^k,A^k,U} \log \left( \frac{P_{X^{n-k},Y^{n-k},X^k,Y^k,A^k,U}}{P_{X^{n-k},Y^{n-k},X^k,Y^k}} \right)
\]

\[
+ \mathbb{E}_{X^k,Y^k,A^k|U} \log \left( \frac{P_{X^k,Y^k}}{P_{X^k,Y^k,A^k,U}} \right) \quad (4.1)
\]

Since $P_{X^{n-k},Y^{n-k},X^k,Y^k,A^k,U} \leq P_{X^{n-k},Y^{n-k},X^k,Y^k}$ the term

\[
\log \left( \frac{P_{X^{n-k},Y^{n-k},X^k,Y^k,A^k,U}}{P_{X^{n-k},Y^{n-k},X^k,Y^k}} \right) \leq 0
\]

Therefore,

\[
(4.1) \leq \mathbb{E}_{X^k,Y^k,A^k|U} \log \left( \frac{P_{X^k,Y^k}}{P_{X^k,Y^k,A^k,U}} \right)
\]

\[
= \mathbb{E}_{X^k,Y^k,A^k|U} \log \left( \frac{1}{P_{A^k,U|X^k,Y^k}} \right) \quad (4.2)
\]

Since $P(A^k = a^k, U|X^k = x^k, Y^k = y^k)$ is a function of only $a^k, x^k, y^k$ (and not of $x^{n-k}, y^{n-k}$) we obtain:

\[
(4.2) = \mathbb{E}_{X^k,Y^k,A^k|U} \log \left( \frac{1}{P(U|X^k,Y^k)} \right) + \mathbb{E}_{X^k,Y^k,A^k|U} \log \left( \frac{1}{P(A^k|X^k,Y^k,U)} \right) \quad (4.3)
\]
In the same way,

\[ (4.3) = E_{X^k, Y^k | U} \log \left( \frac{1}{\Pr(U | X^k, Y^k)} \right) + E_{X^k, Y^k | U} \mathbb{E}_{A^k | X^k, Y^k, U} \log \left( \frac{1}{\Pr(A^k | X^k, Y^k, U)} \right) \]

\[ = \sum_{x^k, y^k \in \text{supp}(P_{X^k, Y^k | U})} \Pr(X^k = x^k, Y^k = y^k | U) \log \left( \frac{1}{\Pr(U | X^k = x^k, Y^k = y^k)} \right) \]

\[ + E_{X^k, Y^k | U} \mathbb{H}(P_{A^k | X^k, Y^k, U}) \]

(4.4)

By the concavity of \( \log(\cdot) \),

\[ (4.4) \leq \log \left( \sum_{x^k, y^k \in \text{supp}(P_{X^k, Y^k | U})} \frac{\Pr(X^k = x^k, Y^k = y^k | U)}{\Pr(U | X^k = x^k, Y^k = y^k)} \right) + E_{X^k, Y^k | U} \mathbb{H}(P_{A^k | X^k, Y^k, U}) \]

\[ = \log \left( \sum_{x^k, y^k \in \text{supp}(P_{X^k, Y^k | U})} \frac{\Pr(X^k = x^k, Y^k = y^k)}{\Pr(U)} \right) + E_{X^k, Y^k | U} \mathbb{H}(P_{A^k | X^k, Y^k, U}) \]

\[ \leq \log \left( \frac{1}{\Pr(U)} \right) + E_{X^k, Y^k | U} \mathbb{H}(P_{A^k | X^k, Y^k, U}) \]

We define \( W \) to be the event that the provers win all the games in the last \( k \) coordinates and define \( E \) to be

\[ \{(a^k, x^k, y^k) \in A^k \times X^k \times Y^k \mid \Pr(A^k = a^k | X^k = x^k, Y^k = y^k) \geq 2^{-\epsilon(n-k)/16}\} \]

The event \( W' \) is defined as \( W \land [(A^k, X^k, Y^k) \in E] \)

**Proposition 4.3.6.** For \( W \) and \( W' \), the events defined above, the following holds:

1. For general games and the event \( W \)

   \[ \mathbb{E}_{X^k, Y^k | W} \mathbb{H}(P_{A^k | X^k, Y^k, W}) \leq k \log s \quad [32],[24] \]

2. For projection games and the event \( W' \)

   \[ \mathbb{E}_{X^k, Y^k | W'} \mathbb{H}(P_{A^k | X^k, Y^k, W'}) \leq \epsilon(n-k)/16 \quad [31] \]
4.3. OUR RESULTS

for general games. We use the trivial bound on the size of the support, namely, for every $x^k, y^k$ we can bound $|\text{supp}(P_{A^k|X^k=x^k,Y^k=y^k,W})| \leq |\text{supp}(P_{A^k})| \leq s^k$ where $s$ is the size of the answers set. Using Fact 2.1.2 we obtain:

$$\mathbb{E}_{X^k,Y^k|W} H(P_{A^k|X^k,Y^k,W}) \leq \mathbb{E}_{X^k,Y^k|W} \log(|\text{supp}(P_{A^k|X^k,Y^k,W})|) \leq \log s^k = k \log s$$

for projection games. Using Fact 2.1.2 we can trivially bound:

$$\mathbb{E}_{X^k,Y^k|W'} H(P_{A^k|X^k,Y^k,W'}) \leq \mathbb{E}_{X^k,Y^k|W'} \log(|\text{supp}(P_{A^k|X^k,Y^k,W'})|)$$

(4.5)

Since for every $x^k, y^k$ and $a^k \in \text{supp}(P_{A^k|X^k=x^k,Y^k=y^k,W'})$,

$$\Pr(A^k = a^k|X^k = x^k, Y^k = y^k) \geq 2^{-\epsilon(n-k)/16},$$

there are at most $2^{(n-k)/16}$ such $a^k$. Hence,

$$(4.5) \leq \mathbb{E}_{X^k,Y^k|W'} \log \left(2^{\epsilon(n-k)/16}\right) = \epsilon(n-k)/16$$

Corollary 4.3.7. For the events $W$, $W'$ the following holds:

1. For general games and the event $W$

$$\mathbb{E}_{i\in[n-k]}\mathbb{E}_{X^k,Y^k,A^k|W} D(P_{X_i,Y_i|X^k,Y^k,A^k,W}\|P_{X_i,Y_i}) \leq \frac{1}{n-k} \left(k \log s - \log(\Pr(W))\right)$$

2. For projection games and the event $W'$

$$\mathbb{E}_{i\in[n-k]}\mathbb{E}_{X^k,Y^k,A^k|W'} D(P_{X_i,Y_i|X^k,Y^k,A^k,W'}\|P_{X_i,Y_i})$$

$$\leq \frac{1}{n-k} \left(\epsilon(n-k)/16 - \log \left(\Pr(W) - 2^{-\epsilon(n-k)/16}\right)\right)$$

(for $z < 0$ we define $\log(z) = -\infty$.)
Chapter 4. Parallel Repetition Theorems for Free Games

Proof. For the general case, fixing $U = W$ in Lemma 4.3.5 and using the bound on $\mathbb{E}_{X^k, Y^k | W} H(P_{A^k | X^k, Y^k, W})$ from Proposition 4.3.6 we obtain:

$$\mathbb{E}_{X^k, Y^k, A^k | W} D\left( P_{X^{n-k}, Y^{n-k} | X^k, Y^k, A^k, W} \parallel P_{X^{n-k}, Y^{n-k}} \right) \leq k \log s - \log(\Pr(W))$$

To complete the proof apply Fact 2.1.4.

For the projection game case, fix $U = W'$ in Lemma 4.3.5 and use the bound on $\mathbb{E}_{X^k, Y^k | W'} H(P_{A^k | X^k, Y^k, W'})$ from Proposition 4.3.6 to obtain:

$$\mathbb{E}_{X^k, Y^k, A^k | W'} D\left( P_{X^{n-k}, Y^{n-k} | X^k, Y^k, A^k, W'} \parallel P_{X^{n-k}, Y^{n-k}} \right) \leq \epsilon(n - k)/16 - \log(\Pr(W'))$$

We bound $\Pr(W')$ in the following way:

$$\Pr(W') = \Pr(W \land [(A^k, X^k, Y^k) \in E]) = \Pr(W) - \Pr(W \land [(A^k, X^k, Y^k) \notin E])$$

We now bound the term $\Pr(W \land [(A^k, X^k, Y^k) \notin E])$. For every game $G$ and strategy $f_a, f_b$, the probability of winning the game played with strategy $f_a, f_b$ is

$$\mathbb{E}_{X,Y} \sum_{b \in B} \Pr(B = b|Y) \sum_{a \in A} \Pr(A = a|X) V(X, Y, a, b).$$

Recall that for every projection game $G$ and every $x \in X, y \in Y, b \in B$ there is only one $a \in A$ for which $V(x, y, a, b) = 1$, this $a$ is $f_{xy}(b)$ (recall that $f_{xy}$ is the projection function, see Definition 4.2.2). Thus for every projection game $G$ and strategy $f_a, f_b$, the probability of winning the game played according to $f_a, f_b$ is:

$$\mathbb{E}_{X,Y} \sum_{(b, f_{xy}(b)) \in B \times A} \Pr(B = b|Y) \Pr(A = a|X).$$

For $x^k, y^k$ we define $f_{x^k, y^k} : B^k \to A^k$ by $[f_{x^k, y^k}(b^k)]_i = f_{x_i, y_i}(b_i)$. We want to bound the probability of winning in the last $k$ coordinates and that $(A^k, X^k, Y^k) \notin E$. Thus, for every $x^k, y^k$ we want to sum $\Pr(B^k = b^k|Y^k = y^k) \Pr(A^k = a^k|X^k = x^k)$, only over
(b^k, f_{x^k, y^k}(b^k)) \in B^k \times A^k$ for which $(f_{x^k, y^k}(b^k), x^k, y^k) \notin E$. Thus

$$\Pr(W \land [(A^k, X^k, Y^k) \notin E])$$

$$= \mathbb{E}_{X^k, Y^k} \sum_{(b^k, f_{X^k, Y^k}(b^k)) \text{ s.t. } (f_{X^k, Y^k}(b^k), X^k, Y^k) \notin E} \Pr(B^k = b^k | Y^k) \Pr(A^k = f_{X^k, Y^k}(b^k) | X^k, Y^k)$$

$$< 2^{-\epsilon(n-k)/16} \quad (4.6)$$

where the last inequality follows since if $(a^k, x^k, y^k) \notin E$ then

$$\Pr(A^k = a^k | X^k = x^k) = \Pr(A^k = a^k | X^k = x^k, Y^k = y^k) < 2^{-\epsilon(n-k)/16}.$$

Thus $\Pr(W') > \Pr(W) - 2^{-\epsilon(n-k)/16}$. We now conclude that

$$\mathbb{E}_{X^k, Y^k, A^k | W'} D \left( P_{X^{n-k}, Y^{n-k} | X^k, Y^k, A^k, W'} \| P_{X^{n-k}, Y^{n-k}} \right) \leq$$

$$\epsilon(n-k)/16 - \log \left( \Pr(W) - 2^{-\epsilon(n-k)/16} \right)$$

The corollary follows by using Fact 2.1.4. \hfill \Box

**Observation 4.3.8.** For any product distribution $P_{\alpha, \beta} = P_{\alpha} \times P_{\beta}$ and any event $\tau$ that is determined only by $\alpha$ (or only by $\beta$) $P_{\alpha, \beta | \tau}$ is a product distribution

$$P_{\alpha, \beta | \tau} = P_{\alpha | \tau} \times P_{\beta | \tau} = P_{\alpha | \tau} \times P_{\beta}$$

(or $P_{\alpha, \beta | \tau} = P_{\alpha} \times P_{\beta | \tau}$)

**Proposition 4.3.9.** For a free game $G$, an event $U$ that is determined by $X^k, Y^k, A^k, B^k$ and for every $x^k, y^k, a^k$ the following holds:

$$P_{X^{n-k} Y^{n-k} | X^k=x^k, Y^k=y^k, A^k=a^k, U} = P_{X^{n-k} | X^k=x^k, Y^k=y^k, A^k=a^k, U} \times P_{Y^{n-k} | X^k=x^k, Y^k=y^k, A^k=a^k, U}$$

That is $P_{X^{n-k} Y^{n-k} | X^k=x^k, Y^k=y^k, A^k=a^k, U}$ is a product distribution.

**Proof.** By applying Observation 4.3.8 three times on the events $X^k = x^k, Y^k = y^k, A^k = a^k$, we obtain that

$$P_{X^{n-k} Y^{n-k} | X^k=x^k, Y^k=y^k, A^k=a^k}$$
is a product distribution. Since after we fixed $x^k, y^k, a^k$, the event $U$ only depends on $B^k$, which is only a function of $Y^{n-k}$, we can apply Observation 4.3.8 one more time to obtain the proposition.

**Corollary 4.3.10.** For a free game $G$, any event $U$ that is determined by $X^k, Y^k, A^k, B^k$ and for every $x^k, y^k, a^k, x, y$ and every $i \in [n-k]$ the following holds:

$$P_{X^{n-k}Y^{n-k} | X^k=x^k, Y^k=y^k, A^k=a^k, U, X_i=x, Y_i=y} = P_{X^{n-k} | X^k=x^k, Y^k=y^k, A^k=a^k, U} \times P_{Y^{n-k} | X^k=x^k, Y^k=y^k, A^k=a^k, U, Y_i=y}$$

**Proof.** From Proposition 4.3.9 we obtain that:

$$P_{X^{n-k}Y^{n-k} | X^k=x^k, Y^k=y^k, A^k=a^k, U}$$

is a product distribution

$$P_{X^{n-k}Y^{n-k} | X^k=x^k, Y^k=y^k, A^k=a^k, U} = P_{X^{n-k} | X^k=x^k, Y^k=y^k, A^k=a^k, U} \times P_{Y^{n-k} | X^k=x^k, Y^k=y^k, A^k=a^k, U}$$

Applying Observation 4.3.8 on the event $X_i = x$ we obtain that

$$P_{X^{n-k}Y^{n-k} | X^k=x^k, Y^k=y^k, A^k=a^k, U, X_i=x} = P_{X^{n-k} | X^k=x^k, Y^k=y^k, A^k=a^k, U, X_i=x} \times P_{Y^{n-k} | X^k=x^k, Y^k=y^k, A^k=a^k, U}$$

Applying Observation 4.3.8 on the event $Y_i = y$ we obtain that

$$P_{X^{n-k}Y^{n-k} | X^k=x^k, Y^k=y^k, A^k=a^k, U, X_i=x, Y_i=y} = P_{X^{n-k} | X^k=x^k, Y^k=y^k, A^k=a^k, U, X_i=x} \times P_{Y^{n-k} | X^k=x^k, Y^k=y^k, A^k=a^k, U, Y_i=y}$$

Recall that for fixed $i \in [n-k]$, we denote $X^m = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_{n-k})$, i.e., $X^{n-k}$ without $X_i$, and similarly, $Y^m = (Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_{n-k})$. 

75
of Lemma 4.3.3 and Lemma 4.3.4. For both $U = W$ and $U = W'$ and for every $x^k, y^k, a^k$ and $i \in [n - k]$, we will use a strategy for the game $G(P_{X^k, Y^k, V^{\otimes n}})$ to obtain a strategy for the game $G(P_{X_i, Y_i | X^k = x^k, Y^k = y^k, A^k = a^k, U, V})$. Fix any strategy, $f_a, f_b$, for the game $G(P_{X^k, Y^k, V^{\otimes n}})$, and apply the following to obtain a strategy for $G(P_{X_i, Y_i | X^k = x^k, Y^k = y^k, A^k = a^k, U, V})$:

Algorithm 4.3.11. Protocol for $G(P_{X_i, Y_i | X^k = x^k, Y^k = y^k, A^k = a^k, U, V})$ for fixed $x^k, y^k, a^k, i$

1. When the game starts, prover 1 receives a question $x$ and prover 2 receives a question $y$ according to $P_{X_i, Y_i | X^k = x^k, Y^k = y^k, A^k = a^k, U}$. Define $X_i = x, Y_i = y$ (the provers will play this game in coordinate $i$).

2. Prover 1 randomly chooses $x^m = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{n-k})$ according to $P_{X^{n-k} | X^k = x^k, Y^k = y^k, A^k = a^k, U, X_i = x}$ and Prover 2 randomly chooses $y^m = (y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_{n-k})$ according to $P_{Y^{n-k} | X^k = x^k, Y^k = y^k, A^k = a^k, U, Y_i = y}$

3. Prover 1 answers $[f_a(x^n)]_i$ and prove 2 answers $[f_b(y^n)]_i$.

Remark 4.3.12. Notice that in step 2, since both events $U = W$ and $U = W'$ are determined by $X^k, Y^k, A^k, B^k$, the joint distribution of $x^m, y^m$ is

$$P_{X^m, Y^m | X^k = x^k, Y^k = y^k, A^k = a^k, X_i = x, Y_i = y, U}$$

which follows from Corollary 4.3.10.

Remark 4.3.13. Notice that since Remark 4.3.12 holds, the probability of winning the game

$$G(P_{X_i, Y_i | X^k = x^k, Y^k = y^k, A^k = a^k, U, V})$$

is exactly

$$\Pr(W_i | X^k = x^k, Y^k = y^k, A^k = a^k, U).$$
Remark 4.3.14. Notice that this is a randomized algorithm. However, it is well known that since any randomized algorithm is a convex combination of deterministic algorithms, there is a deterministic algorithm that achieves the same value as the randomized algorithm. Namely, there is a deterministic protocol for which the probability of winning the game

\[ G(P_{X,Y}|X^k=x^k,Y^k=y^k,A^k=a^k,U,V) \]

is exactly

\[ \Pr(W_i|X^k=x^k,Y^k=y^k,A^k=a^k,U). \]

Using this remark we will think of this algorithm as a deterministic algorithm.

Proof for General Games

By Corollary 4.3.7 for a fixed strategy \( f_a, f_b \) for \( G(P_{X,Y}, V^\otimes n) \),

\[ \mathbb{E}_{i \in [n-k]} \mathbb{E}_{X^k,Y^k,A^k|W} D\left( P_{X_i,Y_i|X^k,Y^k,A^k,W} \| P_{X_i,Y_i} \right) \leq \frac{1}{n-k} (k \log s - \log(\Pr(W))) \]

By the assumption in the lemma, \( \Pr(W) \geq 2^{-\epsilon(n-k)/9+klog s} \). Therefore, it follows that:

\[ \mathbb{E}_{i \in [n-k]} \mathbb{E}_{X^k,Y^k,A^k|W} D\left( P_{X_i,Y_i|X^k,Y^k,A^k,W} \| P_{X_i,Y_i} \right) \leq \epsilon/9 \]

Assume by way of contradiction that for all \( i \in [n-k] \), \( \Pr(W_i|W) > 1 - \epsilon/9 \). Notice that since

\[ \Pr(W_i|W) = \mathbb{E}_{X^k,Y^k,A^k|W} \Pr(W_i|X^k,Y^k,A^k,W), \]

an equivalent assumption is that for all \( i \in [n-k] \),

\[ \mathbb{E}_{X^k,Y^k,A^k|W} \Pr(\neg W_i|X^k,Y^k,A^k,W) < \epsilon/9. \]

By a simple averaging argument, there are \( x^k, y^k, a^k \) and \( i \in [n-k] \) for which both equations hold:

\[ D \left( P_{X_i,Y_i|X^k=x^k,Y^k=y^k,A^k=a^k,W} \| P_{X_i,Y_i} \right) \leq \epsilon/4 \quad (4.7) \]
4.3. OUR RESULTS

\[ \Pr(\neg W_i|X^k = x^k, Y^k = y^k, A^k = a^k, W) < \epsilon/4 \]  \hspace{1cm} (4.8)

For the strategy \( f_a, f_b \) and for \( x^k, y^k, a^k, i \) for which both Equation (4.7) and Equation (4.8) hold consider the protocol suggested in Algorithm 4.3.11. Recall that by Remark 4.3.14 there is a deterministic protocol for which the provers win on coordinate \( i \) with probability

\[ \Pr(W_i|X^k = x^k, Y^k = y^k, A^k = a^k, W). \]

Denote this deterministic protocol by \( h_a, h_b \). For \( h_a, h_b \), denote by \( R \) the set of all questions on which the provers error when playing according to this protocol. By the assumption in Equation (4.8)

\[ P_{X_i,Y_i|X^k=x^k,Y^k=y^k,A^k=a^k,W}(R) < \epsilon/4. \]  \hspace{1cm} (4.9)

Combining Equation (4.9) with Equation (4.7), we can apply Corollary 4.3.2 to obtain \( P_{X_i,Y_i}(R) < \epsilon \). The provers can play \( h_a, h_b \) as a strategy for \( G(P_{X_i,Y_i}, V) \) and error only on questions in \( R \). Since \( P_{X_i,Y_i}(R) < \epsilon \), the value of \( G(P_{X_i,Y_i}, V) > 1 - \epsilon \) and since \( P_{X_i,Y_i} = P_{XY} \), the value of \( G(P_{XY}, V) > 1 - \epsilon \) which is a contradiction.

Proof for Projection Games

The proof is very similar to the general case. From Corollary 4.3.7 we obtain:

\[ \mathbb{E}_{i \in [n-k]} \mathbb{E}_{X^k,Y^k,A^k|W'} D \left( P_{X_i,Y_i|X^k,Y^k,A^k,W'} || P_{X_i,Y_i} \right) \]

\[ \leq \frac{1}{n-k} \left( \epsilon(n-k)/16 - \log \left( \Pr(W) - 2^{-\epsilon(n-k)/16} \right) \right) \]  \hspace{1cm} (4.11)
By the assumption in the lemma, \( \Pr(W) \geq 2^{-\epsilon(n-k)/144} \) thus,

\[
\mathbb{E}_{i \in [n-k]} \mathbb{E}_{X^k,Y^k,A^k|W'} D\left( P_{X_i,Y_i|X^k,Y^k,A^k,W'} \parallel P_{X_i,Y_i} \right)
\leq \epsilon/16 - \frac{1}{n-k} \log \left( 2^{-\epsilon(n-k)/144} - 2^{-\epsilon(n-k)/16} \right)
= \epsilon/16 - \frac{1}{n-k} \log \left( 2^{-\epsilon(n-k)/16} \left( 2^{\epsilon(n-k)/18} - 1 \right) \right)
= \epsilon/16 + \epsilon/16 - \frac{1}{n-k} \log \left( 2^{\epsilon(n-k)/18} - 1 \right)
\leq \epsilon/8
\]

(4.12)

where the last inequality is due to the bound on \( n - k \). Assume by way of contradiction that for all \( i \in [n-k] \), \( \Pr(W_i|W') > 1 - \epsilon/8 \). Notice that since

\[
\Pr(W_i|W') = \mathbb{E}_{X^k,Y^k,A^k|W'} \Pr(W_i|X^k,Y^k,A^k,W')
\]

an equivalent assumption is that for all \( i \in [n-k] \),

\[
\mathbb{E}_{X^k,Y^k,A^k|W'} \Pr(\neg W_i|X^k,Y^k,A^k,W') < \epsilon/8.
\]

By a simple averaging argument, there are \( x^k, y^k, a^k \) and \( i \in [n-k] \) for which both equations hold:

\[
D\left( P_{X_i,Y_i|X^k=Y^k=A^k,W'} \parallel P_{X_i,Y_i} \right) \leq \epsilon/4
\]

(4.13)

\[
\Pr(\neg W_i|X^k = x^k, Y^k = y^k, A^k = a^k, W') < \epsilon/4
\]

(4.14)

For the strategy \( f_a, f_b \), and for \( x^k, y^k, a^k, i \) for which both Equation (4.13) and Equation (4.14) hold consider the protocol suggested in Algorithm 4.3.11. Recall that by Remark 4.3.14 there is a deterministic protocol for which the provers win on coordinate \( i \) with probability

\[
\Pr(W_i|X^k = x^k, Y^k = y^k, A^k = a^k, W')
\]

Denote this deterministic protocol by \( h_a, h_b \). For \( h_a, h_b \), denote by \( R \) the set of all questions on which the provers error when playing according to this protocol. By our
4.3. OUR RESULTS

\[ P_{X_i, Y_i | X^k = x^k, Y^k = y^k, A^k = a^k, W'(R)} < \epsilon/4. \quad (4.15) \]

Combining Equation (4.15) with Equation (4.13), we can apply Corollary 4.3.2 to obtain

\[ P_{X_i, Y_i}(R) < \epsilon. \]

The provers can play \( h_a, h_b \) as a strategy for \( G(P_{X_i, Y_i, V}) \) and error only on questions in \( R \). Since \( P_{X_i, Y_i}(R) < \epsilon \), the value of \( G(P_{X_i, Y_i, V}) > 1 - \epsilon \). Since \( P_{X_i, Y_i} = P_{XY} \) the value of \( G(P_{XY}, V) > 1 - \epsilon \) which is a contradiction.

We showed that there is \( i \in [n - k] \) for which

\[ \Pr(W_i | W') \leq 1 - \epsilon/8 \]

but we need to show that there is \( i \in [n - k] \) for which \( \Pr(W_i | W) \leq 1 - \epsilon/9 \). This is done in the following way: Since \( W' \subseteq W \)

\[ \Pr(W_i | W) = \Pr(W_i | W') \Pr(W' | W) + \Pr(W_i | \neg W') \Pr(\neg W' | W) \leq \Pr(W_i | W') + \Pr(\neg W' | W). \]

Thus for all \( i \in [n - k], \)

\[ \Pr(W_i | W) \leq \Pr(W_i | W') + \Pr((A^k, X^k, Y^k) \notin E | W). \]

Since \( \Pr((A^k, X^k, Y^k) \notin E | W) = \Pr(W \land [(A^k, X^k, Y^k) \notin E]) / \Pr(W) \) we can use the bound in Equation (4.6), \( \Pr(W \land [(A^k, X^k, Y^k) \notin E]) < 2^{-\epsilon(n-k)/16} \) and obtain that

\[ \Pr(W_i | W) \leq \Pr(W_i | W') + 2^{-\epsilon(n-k)/16} / \Pr(W). \]

Therefore:

\[ \Pr(W_i | W) \leq 1 - \epsilon/8 + 2^{-\epsilon(n-k)/16} / 2^{-\epsilon(n-k)/144} \]

\[ \leq 1 - \epsilon/8 + 2^{-\epsilon(n-k)/18} \]

\[ \leq 1 - \epsilon/9 \]

where the last inequality follows from the bound on \( n - k \) \( \square \)
Proof Of Theorem 8. We first show by induction that for every \( k \leq \frac{en}{18 \log s + 3} \) there is a set \( T \subseteq [n] \) of \( k \) coordinates (\(|T| = k\)) for which \( \Pr(W) \leq (1 - \epsilon/9)^k \) where \( W \) is the event of winning on all the coordinates in \( T \). For \( k = 0 \) the statement trivially holds. Assume by induction that there is a set \( T \) of size \( k \) for which \( \Pr(W) \leq (1 - \epsilon/9)^k \). If \( \Pr(W) \leq (1 - \epsilon/9)^{k+1} \) then we are done. Otherwise
\[
\Pr(W) > (1 - \epsilon/9)^{k+1} \geq 2^{-\epsilon(k+1)/4.5}
\]
where we used the inequality \((1 - x) \geq 2^{-2x}\) for \( 0 \leq x \leq 1/2 \). In order to use Lemma 4.3.3 we need to make sure that \( \Pr(W) \geq 2^{-\epsilon(n-k)/9+k \log s} \). It is enough to show that
\[
2^{-\epsilon(k+1)/4.5} \geq 2^{-\epsilon(n-k)/9+k \log s}
\]
or alternatively,
\[
\epsilon(k + 1)/4.5 \leq \epsilon(n - k)/9 - k \log s
\]
After rearranging we obtain
\[
k \leq \frac{en - 2\epsilon}{9 \log s + 3\epsilon}.
\]
For \( n > 2 \) and \( \epsilon \leq 1/2 \) it is enough that\(^2\)
\[
k \leq \frac{en}{18 \log s + 3}.
\]
Thus, for \( k \leq \frac{en}{18 \log s + 3} \) we can apply Lemma 4.3.3 to obtain that there is \( i \notin T \) for which \( \Pr(W_i|W) \leq 1 - \epsilon/9 \) therefore,
\[
\Pr(W_i \land W) = \Pr(W) \cdot \Pr(W_i|W) \leq (1 - \epsilon/9)^k \land (1 - \epsilon/9) = (1 - \epsilon/9)^{k+1}
\]
To complete the proof, set \( k = \frac{en}{18 \log s + 3} \) then as we showed, there is a set \( T \subseteq [n], |T| = k \) for which:
\[
\Pr(W_1 \land \ldots \land W_n) \leq \Pr(\bigwedge_{i \in T} W_i) \leq (1 - \epsilon/9)^{en/(18 \log s + 3)} \leq (1 - \epsilon^2/9)^n/(18 \log s + 3).
\]
where the last inequality follows by the use of the inequality \((1 - x)^y \leq 1 - xy\) for every \( 0 \leq y \leq 1 \) and \( x \leq 1 \).\(\square\)

\(^2\)We may assume that \( n > 2 \) since for \( n \leq 2 \) the theorem trivially holds. We also assume that the game is not trivial, i.e., the value of the game is not 0 or 1, thus \( s > 1 \).
4.3. Proof of Main Results

Of Theorem 9. For the case of \( n \geq (50/\epsilon) \log(8/\epsilon) \), the proof is very similar to the last theorem: We first show by induction, for every \( k \leq (n/33) - 1 \) there is a set \( T \subseteq [n] \) of \( k \) coordinates \(|T| = k\) for which \( \Pr(W) \leq (1 - \epsilon/9)^k \) where the event \( W \) is winning on all the coordinates in \( T \). For \( k = 0 \) the statement trivially holds. Assume by induction that there is a set \( T \) of size \( k \) for which \( \Pr(W) \leq (1 - \epsilon/9)^k \). If \( \Pr(W) \leq (1 - \epsilon/9)^{k+1} \) then we are done, else

\[
\Pr(W) \geq (1 - \epsilon/9)^{k+1} \geq 2^{-\epsilon(k+1)/4.5}.
\]

In order to use Lemma 4.3.4 we need to make sure that

\[
\Pr(W) \geq 2^{-\epsilon(n-k)/144}
\]

and that

\[
n - k \geq (48/\epsilon) \log(8/\epsilon)
\]

Since \( k \leq (n/33) - 1 \),

\[
\text{if } \Pr(W) \geq 2^{-\epsilon(k+1)/4.5} \text{ then } \Pr(W) \geq 2^{-\epsilon(n-k)/144}
\]

Since \( k \leq (n/33) - 1 \) then \( n - k \geq 32n/33 + 1 \). Since \( n \geq 50/\epsilon \log(8/\epsilon) \) then

\[
32n/33 + 1 \geq (48/\epsilon) \log(8/\epsilon) + 1.
\]

Therefore,

\[
n - k \geq (48/\epsilon) \log(8/\epsilon)
\]

Now we can apply Lemma 4.3.4 to obtain that there is \( i \notin T \) for which \( \Pr(W_i|W) \leq 1 - \epsilon/9 \). Therefore,

\[
\Pr(W_i \land W) = \Pr(W) \cdot \Pr(W_i|W) \leq (1 - \epsilon/9)^k (1 - \epsilon/9) = (1 - \epsilon/9)^{k+1}
\]

For \( k = (n/33) - 1 \) there is a set \( T \subseteq [n], |T| = k \) for which:

\[
\Pr(W_1 \land \ldots \land W_n) \leq \Pr(\bigwedge_{i \in T} W_i) \leq (1 - \epsilon/9)^{(n/33) - 1}
\]
For the case of $n < (50/\epsilon) \log(8/\epsilon)$, as suggested in [31], it can be shown that if the theorem was false for small $n$ it would not hold for big $n$. If there was a strategy with success probability greater than $(1 - \epsilon/9)^{(n/33)-1}$ then for the same game played on $m \cdot n$ coordinates the success probability was at least $(1 - \epsilon/9)^{m((n/33)-1)}$ and for large enough $m$, this yield a contradiction.
Chapter 5

Parallel Repetition Theorems for k-Provers in a No-Signaling Model

5.1 Introduction

The parallel repetition theorem states that for any two provers one round game with value at most $1 - \epsilon$ (for $\epsilon < 1/2$), the value of the game repeated $n$ times in parallel is at most $(1 - \epsilon^3)^{\Omega(n/\log s)}$ where $s$ is the size of the answers set. It is not known how the value of the game decreases when there are three or more players. In this chapter we address the problem of the error decrease of parallel repetition game for $k$-provers where $k > 2$. We consider a special case of the No-Signaling model and show that the error of the parallel repetition of $k$ provers one round game, for $k > 2$, in this model, decreases exponentially depending only on the error of the original game and on the number of repetitions.

In a $k$ provers one round game there are $k$ provers and a verifier. The verifier selects randomly $(x_1, \ldots, x_k) \in X_1 \times \cdots \times X_k$, a question for each prover, according to some distribution $P_{X_1, \ldots, X_k}$ where $X_i$ is the questions set of prover $i$ where $i \in 1, \ldots, k$. Each prover knows only the question addressed to her, prover 1 knows only $x_1$ and prover $i$
knows only $x_i$. The provers cannot communicate during the transaction. The provers send their answers to the verifier, $a_i \in A_i$ where $A_i$ is the answers set of prover $i$. The verifier evaluates an acceptance predicate $V(x_1, \ldots, x_k, a_1, \ldots, a_k)$ and accepts or rejects based on the outcome of the predicate. The acceptance predicate as well as the distribution of the questions are known in advance to the provers. The provers answer the questions according to a strategy. The strategy of the provers is also called a protocol. The type of the protocol determines the type of model. In the *classic model* the provers’ strategy is a $k$ tuple of functions, one for each prover, where each function is from her questions to her answers, i.e., for all $i$, $f_{a_i}: X_i \rightarrow A_i$. We also call this model $\text{MIP}(k, 1)$. In the *No-Signaling model* the provers’ strategy is a $k$ tuple of functions, one for each prover, where each function is a random function from the questions of all the provers to her answers but in a way that does not reveal any information about the other provers’ questions. We denote this model by $\text{MIP}^{\triangledown}(k, 1)$. The no signaling condition ensures that the answer of each prover given its question is independent of the questions of the other provers (but not of the other provers’ answers). This definition is the natural generalization of the no-signaling model of two provers one round game to no signaling model of $k$ provers one round game. However, in this chapter we generalize the no-signaling model in a different way. In this variation, no $k-1$ provers together can signal to the remaining prover. In the standard model one also demands that no prover can signal to any composite system of at most $k-1$ provers. To get a sense of such games, let us consider a generalization of the CHSH [11], [35] game for 3 players. Let us denote the three provers by Alice, Bob and Charlie. The verifier chooses uniformly at random three questions $x, y, z \in \{0, 1\}$ and sends $x$ to Alice, $y$ to Bob and $z$ to Charlie. The provers sends their answers, denoted by $a, b, c$ and the verifier accepts if and only if $x \land y \land z = a \oplus b \oplus c$. We will consider the following strategy: if $x \land y \land z = 0$ then $(a, b, c)$ is distributed uniformly over $\{(0, 0, 0), (1, 0, 1), (1, 1, 0), (0, 1, 1)\}$. If $x \land y \land z = 1$ then $(a, b, c)$ is distributed uniformly over $\{(1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 1, 1)\}$. This is a no-
Chapter 5. Parallel Repetition Theorems for k-Provers in a No-Signaling Model

signaling strategy since for every question of Alice, knowing her answer to that question does not reveal any information about Bob’s or Charlie’s questions. Similar argument apply to Bob and Charlie.

The value of the game is the maximum of the probability that the verifier accepts, where the maximum is taken over all the provers strategies. More formally, the value of the game is:

$$\max_{f_{a_1}, \ldots, f_{a_k}} \mathbb{E}_{x_1, \ldots, x_k} [V(x_1, \ldots, x_k, f_{a_1}(x_1), \ldots, f_{b}(x_k))]$$

where the expectation is taken with respect to the distribution $P_{X_1, \ldots, X_k}$. We denote the value of the game in the classical model by $w(G)$ and the value of the game in the No-Signaling model by $\omega^\diamondsuit(G)$.

Roughly speaking, the \textit{n-fold parallel repetition} of the game, denoted by $(G^\otimes n)$, is playing the game $n$ times in parallel, i.e. the verifier sends $n$ questions to each prover and receives $n$ answers from each prover. The verifier evaluates the acceptance predicate on each game and accepts if and only if all the predicates were true. Obviously $\omega^\diamondsuit(G^\otimes n) \leq \omega^\diamondsuit(G)$ but one might expect that $\omega^\diamondsuit(G^\otimes n) = (\omega^\diamondsuit(G))^n$. However, although the verifier treats each game of the $n$ games independently, the provers may not. The answer of each question addressed to a prover may depend on all the questions addressed to that prover. There are examples for the MIP(2, 1) model [22], [20], [32], [21], [24], and for the \textit{MIP}^{\diamondsuit}(2, 1) model [24] for which the $\omega^\diamondsuit(G^\otimes n) = (\omega^\diamondsuit(G))^n$ for small $n$. Raz provided an example [16] for MIP(k, 1) where $\omega^\diamondsuit(G^\otimes k) = (\omega^\diamondsuit(G))$.

Another model of $k$-provers one round game is the \textit{quantum model}, denoted by $\text{MIP}^*(k, 1)$, in which the provers have a joint quantum state and so the function of each prover is from the questions of all the provers to her answers but only for such functions that can be implemented by an entangled quantum state.

Clearly, the value of a game in the MIP(k, 1) model is less or equal to the value of the game in the $\omega^*(G)$ model which is less or equal to the value of the game in $\text{MIP}^\diamondsuit(k, 1)$ model. However, this bound does not imply a bound for the parallel repetition of the
MIP($k, 1$) model since there are games for which the value of the game in the MIP($k, 1$) model is strictly less than 1 but the value of the game in $MIP^\diamondsuit(k, 1)$ model is 1, therefore implying only the trivial bound. If the value of the game in $MIP^\diamondsuit(k, 1)$ model is strictly less than 1 then we obtain a non-trivial bound on the value of the game for the MIP($k, 1$) model.  

Many fundamental questions related to the MIP(2, 1) model have been answered by now, but there was no known upper bounds on MIP($k, 1$) for $k > 2$ in any model\(^2\). However, it was conjectured that the error decreases exponentially. Although this model of no-signaling is not the natural no-signaling model, it allows us to obtain a first upper bound result for the MIP($k, 1$) for $k > 2$ for special types of games\(^3\) and it raises explicitly the questions for what models one can look at more than 2 provers.

### 5.1.1 Related Work

A series of papers deals with the nature of the error decrease of parallel repetition game [10], [15], [18], [32], [24], [31]. The breakthrough result was done by Raz’s [32]. In his celebrated result, Raz proved that the error of MIP(2, 1) decreases exponentially depending also on the size of the answer set. Holenstein [24] simplified this result and revealed new insights on the nature of the problem. Holenstein also improved the constants to give a tighter bound on the error of the $n$ parallel repetition. Holenstein also proved that in the $MIP^\diamondsuit(2, 1)$ model, the error decreases exponentially and does not depend on the size of the answer set.

---

1It decreases exponentially and does not depend on the size of the answers support.

2This is not completely accurate. The paper [37] states its results only for two provers, but its proof works without change for any number of provers. I would like to thank anonymous referee 1 for this comment.

3Namely, games where the value of the game played according to a no-signaling protocol is strictly less than 1.
5.1.2 Our results

A 3-provers one round No-Signaling game

\[ G = (X, Y, Z, A, B, C, P_{XYZ}, Q) \]

is an object consisting on three finite question sets \( X, Y, Z \), a probability measure \( P_{XYZ} \) on \( XYZ \)

\[ P_{XYZ} : X \times Y \times Z \to \mathbb{R}^+, \]

answer sets \( A, B, C \) and an acceptance predicate

\[ Q : X \times Y \times Z \times A \times B \times C \to \{0, 1\}. \]

A No-Signaling protocol is a set of three No-Signaling functions: the function

\[ p_1 : X \times Y \times Z \times R \to A, \]

the function

\[ p_2 : X \times Y \times Z \times R \to B \]

and the function:

\[ p_3 : X \times Y \times Z \times R \to C. \]

A function

\[ p_1 : X \times Y \times Z \times R \to A \]

is a No-Signaling function if:

\[ \forall x \in X, a \in A, y, y' \in Y, z, z' \in Z \quad \text{Pr}_R[p_1(x, y, z, R) = a] = \text{Pr}_R[p_1(x, y', z', R) = a]. \]

For \( k \) provers, a function \( p_i \) is no-signaling if:

\[ \forall x \in X_i, a \in A_i, x_j, x'_j \in X_j \quad \text{for } j \neq i \quad \text{Pr}_R[p_i(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_k, R) = a] = \text{Pr}_R[p_i(x'_1, \ldots, x'_{i-1}, x, x'_{i+1}, \ldots, x'_k, R) = a]. \]
Definition 5.1.1 (value of the game). The value of the game is defined as:

\[ \omega^{\diamond}(G) = \max_{p_1, p_2, p_3} \mathbb{E}_{X,Y,Z,R}[Q(X,Y,Z,p_1(X,Y,Z,R),p_2(X,Y,Z,R),p_3(X,Y,Z,R))] \]

where the expectation is taken with respect to \( P_{X,Y,Z} \) and the maximum is taken over all No-Signaling protocols.

The \( n \)-fold parallel repetition 3-provers game \( G^{\otimes n} \) consists of the sets \( X^n, Y^n, Z^n \), a probability measure on those sets \( P^{\otimes n}_{X^n,Y^n,Z^n} \) where

\[ P^{\otimes n}_{X^n,Y^n,Z^n}(x^n, y^n, z^n) = \prod_{i=1}^n P_{X,Y,Z}(x_i, y_i, z_i) \]

for \( x^n \in X^n, y^n \in Y^n, z^n \in Z^n \). It also consists of the sets \( A^n, B^n, C^n \) and an accepting predicate \( Q^{\otimes n} \) where

\[ Q^{\otimes n}(x^n, y^n, z^n, a^n, b^n, c^n) = \bigwedge_{i=1}^n Q(x_i, y_i, z_i, a_i, b_i, c_i) \]

for \( x^n \in X^n, y^n \in Y^n, z^n \in Z^n, a^n \in A^n, b^n \in B^n, c^n \in C^n \). The \( n \)-fold parallel repetition for \( k \)-provers game (for every \( k > 3 \)) is defined in the obvious way and therefore omitted.

For a game \( G^n \) and a strategy \((p_1, p_2, p_3)\) we define:

\[ P^{\otimes n}_{X^n,Y^n,Z^n,A^n,B^n,C^n}(x^n, y^n, z^n, a^n, b^n, c^n) \triangleq P^{\otimes n}_{X^n,Y^n,Z^n}(x^n, y^n, z^n) \cdot \Pr[p_1(x^n) = a^n \land p_2(y^n) = b^n \land p_3(y^n) = c^n] \]

We can now present our theorem:

**Theorem 10.** For every \( k \geq 2 \), every positive integer \( n \), all games

\[ G = (X,Y,Z,A,B,C,P_{X,Y,Z},Q), \]

played in the MIP^{\diamond}(k, 1) model satisfy:

\[ \omega^{\diamond}(G^{\otimes n}) \leq \left(1 - \frac{(1 - \omega^{\diamond}(G))^2}{100(1 + 4k)^2}\right)^n. \]

Holenstein [24] proved the theorem for \( k = 2 \). For simplicity, we will prove the theorem for \( k = 3 \) and generalize it for \( k > 3 \) in Appendix 7.4.
5.2 Preliminaries

We denote an \( n \)-dimensional vector by superscripts \( n \), e.g., \( A^n = (A_1, \ldots, A_n) \) where \( A_i \) is its \( i^{th} \) entry and \( A^{-i} = (A_1, \ldots, A_{i-1}, A_{i+1}, \ldots, A_n) \). The statistical difference between two probability distributions \( P \) and \( Q \), defined over the same sample space \( \Omega \), is \( \|P - Q\|_1 = \frac{1}{2} \sum_{x \in \Omega} |P(x) - Q(x)| \).

**Definition 5.2.1 (Divergence).** We define the Kullback-Leibler divergence, also called the informational divergence or simply divergence. Let \( P \) and \( Q \) be two probability measures defined on the same sample space, \( \Omega \). The divergence of \( P \) with respect to \( Q \) is:

\[
D(P \parallel Q) = \sum_{x \in \Omega} P(x) \log \frac{P(x)}{Q(x)}
\]

where \( 0 \log 0 \) is defined to be 0 and \( p \log \frac{p}{0} \) where \( p \neq 0 \) is defined to be \( \infty \).

Vaguely speaking, we could think of the divergence as a way to measure the information we gained by knowing that a random variable is distributed according to \( P \) rather than \( Q \). This indicates how far \( Q \) is from \( P \), if we don't gain much information then the two distributions are very close in some sense.

5.3 Proof sketch

We closely follow Holenstein’s proof [24] and generalize it for three provers one round No-Signaling game. Furthermore, in Appendix 7.4, we generalize the result for every \( k \geq 3 \) provers.

Fixing a No-Signaling strategy \( (p_1, p_2, p_3) \), we define \( W_i \) to be the probability of winning game \( i \). Using this notation the parallel repetition is an upper bound for \( \Pr[W_1 \land W_2 \ldots \land W_n] \). Since

\[
\Pr[W_1 \land \ldots \land W_n] = \Pr[W_{i_1}] \Pr[W_{i_2}|W_{i_1}] \cdots \Pr[W_{i_n}|W_{i_1}W_{i_2} \cdots W_{i_{n-1}}]
\]
we will upper bound $\Pr[W_{i_m} | W_{i_1} \ldots W_{i_{m-1}}]$ for every $m \leq n^\delta$ (where $\delta$ is just some small constant).

We will show that conditioning on the event of winning all $m - 1$ games, i.e., conditioning on $W_{i_1} \ldots W_{i_{m-1}}$, there is a game $i_m$ on which the probability of winning that game is at most $\omega^{\lozenge}(G) + \epsilon$. The way to do so, is by showing that given a No-Signaling strategy for the game $G^{\otimes n}$ and conditioned on winning $m - 1$ games, the provers can use this strategy to obtain a No-Signaling strategy for $G$. To obtain a strategy for $G$ we will need two lemmas. The first lemma, Lemma 5.4.2 shows that fixing a strategy and conditioning on the event of winning $m - 1$ games, there is a coordinate $i_m$ for which $P^\otimes n_{X^nY^nZ^n|W_{i_1} \ldots W_{i_{m-2}}}(x^n, y^n, z^n)$ projected on the $i_m$th coordinate is similar to $P_{XYZ}$. This lemma is Claim 5.1 in Raz’s paper [32] and Lemma 5 in Holenstein’s paper [24]. The second lemma, Lemma 5.4.4 shows that we can obtain a No-Signaling strategy to play the $i_m$th coordinate from a No-Signaling strategy for $G^{\otimes n}$ after conditioning on the event $W_{i_1} \land \ldots \land W_{i_{m-1}}$ (after conditioning we may not obtain a No-Signaling function but we show that there is a No-Signaling function which is very close to it.). This is a generalization of Lemma 23 in Holenstein’s paper [24]. We then conclude that the provers cannot win the game $i_m$ with probability greater than $\omega^{\lozenge}(G) + \epsilon$.

5.4 Technical Lemmas

**Lemma 5.4.1** (RazHolenstein). Let $P_{U^\ell} = P_{U_1} P_{U_2} \cdots P_{U_\ell}$ be a probability distribution over $U^\ell$ (i.e., $P_{U^\ell}$ is a product distribution over $U_1, \ldots U_\ell$) and let $W$ be some event, then the following holds:

$$\sum_{i=1}^\ell \| P_{U_i|W} - P_{U_i} \| \leq \sqrt{\ell \cdot \log \frac{1}{\Pr[W]}}$$

For completeness, we include the proof of the lemma in Appendix 7.2.

We need the following corollary of Lemma 5.4.1.
Corollary 5.4.2. Let $P_{TU} = P_T P_{U_1|T} P_{U_2|T} \cdots P_{U_\ell|T}$ and let $W$ be some event, then the following holds:

$$\sum_{i=1}^\ell \|P_{TU_i|W} - P_{T|W} P_{U_i|T}\| \leq \sqrt{\ell \cdot \log \frac{1}{\Pr[W]}}$$

For the proof see Appendix 7.3.

For the second lemma we will need to use the following proposition.

Proposition 5.4.3. Let $P_{RST}$ be a probability distribution over $RST$ and let $\Pi_R$ be some distribution over $R$. There exists a distribution $Q_{RST}$ satisfying the following:

\begin{align}
\|Q_{RST} - P_{RST}\| &\leq \|P_R - \Pi_R\| \\
\|Q_R - \Pi_R\| & = 0 \\
\|Q_S - P_S\| & = 0 \\
\|Q_T - P_T\| & = 0
\end{align}

The proof is following closely Holenstein’s [24] proof of Lemma 22.

Proof. We show a finite process for changing $P_{RST}$ into $Q_{RST}$ with the desired properties. Let $H = \{r \mid P_R(r) > \Pi_R(r)\}$ and let $L = \{r \mid P_R(r) < \Pi_R(r)\}$. Until $|H| \neq 0$, fix some $r \in H$ and $r' \in L$ and fix $s, t$ such that $P_{RST}(r, s, t) > 0$. Define,

$$\alpha := \min\{P_{RST}(r, s, t), |P_R(r) - \Pi_R(r)|, |P_R(r') - \Pi_R(r')|\}$$

and set $P_{RST}(r, s, t) := P_{RST}(r, s, t) - \alpha$ and $P_{RST}(r', s, t) := P_{RST}(r', s, t) + \alpha$. The process only decreases $\|P_R - \Pi_R\|$ and therefore Condition 5.1 holds. By definition, we continue the process until Condition 5.2 holds. Since in every iteration, the process decreases some fixed $(r, s, t)$ by $\alpha$ and increases $(r', s, t)$ by $\alpha$, for the same $s, t$ then Condition 5.3 and Condition 5.4 are satisfied. 

Lemma 5.4.4. Let $P_{XYZRST}$ be a probability distribution over $XYZRST$ and let $\Pi_{XYZ}$
be some distribution over \( XYZ \). If

\[
\|\hat{P}_{XYZ}P_{R|X} - P_{XYZR}\| \leq \epsilon_1 \tag{5.5}
\]
\[
\|\hat{P}_{XYZ}P_{S|Y} - P_{XYZS}\| \leq \epsilon_2 \tag{5.6}
\]
\[
\|\hat{P}_{XYZ}P_{T|Z} - P_{XYZT}\| \leq \epsilon_3 \tag{5.7}
\]

then, there exists a conditional distribution \( Q_{RST|X=x,Y=y,Z=z} \) satisfying the following:

\[
\|\hat{P}_{XYZ}Q_{RST|XYZ} - P_{XYZRST}\| \leq \min\{\epsilon_1, \epsilon_2, \epsilon_3\} + 2(\epsilon_1 + \epsilon_2 + \epsilon_3) \tag{5.8}
\]
\[
\|Q_{R|X=x,Y=y,Z=z} - Q_{R|X=x}\| = 0 \tag{5.9}
\]
\[
\|Q_{S|X=x,Y=y,Z=z} - Q_{S|Y=y}\| = 0 \tag{5.10}
\]
\[
\|Q_{T|X=x,Y=y,Z=z} - Q_{T|Z=z}\| = 0 \tag{5.11}
\]

Proof. For fixed \( x, y, z \) we apply Lemma 5.4.3 three times. We first apply the lemma on \( P_{RST|X=x,Y=y,Z=z} \) and \( P_{R|X=x} \) to obtain \( \hat{Q}_{RST|X=x,Y=y,Z=z} \) such that

\[
\|\hat{Q}_{RST|X=x,Y=y,Z=z} - P_{RST|X=x,Y=y,Z=z}\| \leq \|P_{R|X=x,Y=y,Z=z} - P_{R|X=x}\|
\]
\[
\|\hat{Q}_{R|X=x,Y=y,Z=z} - P_{R|X=x}\| = 0
\]
\[
\|\hat{Q}_{S|X=x,Y=y,Z=z} - P_{S|X=x,Y=y,Z=z}\| = 0
\]
\[
\|\hat{Q}_{T|X=x,Y=y,Z=z} - P_{T|X=x,Y=y,Z=z}\| = 0
\]

Applying Lemma 5.4.3 on \( \hat{Q}_{RST|X=x,Y=y,Z=z} \) and \( P_{S|Y=y} \) and combining the previous result we obtain \( \hat{Q}_{RST|X=x,Y=y,Z=z} \) such that

\[
\|\hat{Q}_{RST|X=x,Y=y,Z=z} - \hat{Q}_{RST|X=x,Y=y,Z=z}\| \leq \|P_{S|X=x,Y=y,Z=z} - P_{S|Y=y}\|
\]
\[
\|\hat{Q}_{R|X=x,Y=y,Z=z} - P_{R|X=x}\| = 0
\]
\[
\|\hat{Q}_{S|X=x,Y=y,Z=z} - P_{S|Y=y}\| = 0
\]
\[
\|\hat{Q}_{T|X=x,Y=y,Z=z} - \hat{Q}_{T|X=x,Y=y,Z=z}\| = 0
\]

By applying Lemma 5.4.3 again, this time on

\[
\hat{Q}_{RST|X=x,Y=y,Z=z}
\]
Chapter 5. Parallel Repetition Theorems for k-Provers in a No-Signaling Model

and $\tilde{Q}_{T|Z=z}$, we obtain,

$$Q_{RST|X=x,Y=y,Z=z}$$

such that

$$\|Q_{RST|X=x,Y=y,Z=z} - \tilde{Q}_{RST|X=x,Y=y,Z=z}\| \leq \|P_{T|X=x,Y=y,Z=z} - P_{T|Z=z}\|$$

$$\|Q_{R|X=x,Y=y,Z=z} - P_{R|X=x}\| = 0$$

$$\|Q_{S|X=x,Y=y,Z=z} - P_{S|Y=y}\| = 0$$

$$\|Q_{T|X=x,Y=y,Z=z} - P_{T|Z=z}\| = 0$$

From Condition 5.5, Condition 5.6 and Condition 5.7 we obtain:

$$\|P_{XYZ} - \bar{P}_{XYZ}\| \leq \min\{\epsilon_1, \epsilon_2, \epsilon_3\}$$

Hence,

$$\|P_{XYZ}Q_{RST|XYZ} - P_{XYZRST}\|$$

$$\leq \min\{\epsilon_1, \epsilon_2, \epsilon_3\} + \|P_{XYZ}Q_{RST|XYZ} - P_{XYZRST}\|$$

$$= \min\{\epsilon_1, \epsilon_2, \epsilon_3\} + \sum_{xyz} P_{XYZ}(x, y, z)\|Q_{RST|X=x,Y=y,Z=z} - P_{RST|X=x,Y=y,Z=z}\|$$

$$\leq \min\{\epsilon_1, \epsilon_2, \epsilon_3\} + \sum_{xyz} P_{XYZ}(x, y, z)\cdot$$

$$\left(\|P_{R|X=x,Y=y,Z=z} - P_{R|X=x}\| + \|P_{S|X=x,Y=y,Z=z} - P_{S|Y=y}\| + \|P_{T|X=x,Y=y,Z=z} - P_{T|Z=z}\|\right)$$

$$\leq \min\{\epsilon_1, \epsilon_2, \epsilon_3\} + \|P_{XYZR} - P_{XYZP_{R|X}}\| + \|P_{XYZS} - P_{XYZP_{S|Y}}\|$$

$$+ \|P_{XYZS} - P_{XYZP_{T|Z}}\|$$

$$\leq \min\{\epsilon_1, \epsilon_2, \epsilon_3\} + 2(\epsilon_1 + \epsilon_2 + \epsilon_3)$$

5.5 Parallel Repetition Theorem

Proposition 5.5.1. For every game $G = (X, Y, Z, A, B, C, P_{XYZ}, Q)$ fix any No-Signaling strategy for the $n$-fold parallel repetition game, $G^{\otimes n}$, and let $W = W_{i_1} \land \ldots \land W_{i_{m-1}}$ be
5.5. PARALLEL REPETITION THEOREM

the event of winning on some fixed \(i_1, \ldots, i_{m-1}\) coordinates with respect to that strategy. There exists a coordinate \(i_m\) such that

\[
\Pr[W_{i_m}|W] \leq \omega^\diamond(G) + 13 \sqrt{\frac{1}{n-m} \log \left( \frac{1}{\Pr[W]} \right)}
\]

Proof. Assume without loss of generality that \(i_1, \ldots, i_{m-1} \in \{n-m+1, \ldots, n\}\) (the last \(m\) coordinates). Let \(T = (X^n, A^n)\) and \(U^{n-m} = (Y^{n-m}, Z^{n-m})\) then from Lemma 5.4.2 we obtain:

\[
\sum_{i=1}^{n-m} \|P_{X^nA^nZ_i}|W - P_{X^nA^n}|W P_{Y_iZ_i|X^nA^n}\| \leq \sqrt{(n-m) \log \frac{1}{Pr[W]}}.
\]

Since the strategy is No-Signaling:

\[
\sum_{i} \|P_{X^nA^nY_iZ_i}|W - P_{X^nA^n}|W P_{Z_iY_i|X_i}\| \leq \sqrt{(n-m) \log \frac{1}{Pr[W]}}. \tag{5.12}
\]

This can be written as:

\[
\sum_{i} \|P_{X^nA^nY_iZ_i}|W - P_{X_i}|W P_{X^{-i}A^n|X_i,W}P_{Z_iY_i|X_i}\| \leq \sqrt{(n-m) \log \frac{1}{Pr[W]}}. \tag{5.13}
\]

Applying Lemma 5.4.1 with \(U^{n-m} = X^{n-m}\) we obtain:

\[
\sum_{i} \|P_{X_i}|W - P_{X_i}\| \leq \sqrt{(n-m) \log \frac{1}{Pr[W]}}.
\]

Combining Equation 5.12 and Equation 5.13 yields:

\[
\sum_{i} \|P_{X^nA^nY_iZ_i}|W - P_{X_iP_{X^{-i}A^n|X_i,W}P_{Z_iY_i|X_i}}\| \leq 2 \sqrt{(n-m) \log \frac{1}{Pr[W]}}.
\]

This can be written as:

\[
\sum_{i} \|P_{X^nA^nY_iZ_i}|W - P_{X_iY_iZ_iP_{X^{-i}A^n|X_i,W}}\| \leq 2 \sqrt{(n-m) \log \frac{1}{Pr[W]}}.
\]

Since \(P_{X_iY_iZ_i} = P_{XYZ}\),

\[
\sum_{i} \|P_{X^nA^nY_iZ_i}|W - P_{XYZP_X^{-i}A^n|X_i,W}\| \leq 2 \sqrt{(n-m) \log \frac{1}{Pr[W]}}.
\]
Similar argument for $T = (Y^n, B^n)$ and $U^{n-m} = (X^{n-m}, Z^{n-m})$ yields:
\[
\sum_{i}^{n-m} ||P_{Y^nB^nX_iZ_i|W} - P_{XYZP_{Y-i|Y_i}}|| \leq 2(n - m) \log \frac{1}{\Pr[W]}
\]

And for $T = (Z^n, C^n)$ and $U^{n-m} = (X^{n-m}, Y^{n-m})$ yields:
\[
\sum_{i}^{n-m} ||P_{Z^nC^nX_i|W} - P_{XYZP_{Z-i|Z_i}}|| \leq 2(n - m) \log \frac{1}{\Pr[W]}
\]

By Lemma 5.4.4, there exists a distribution
\[
Q_{X^{-i}A^nY^{-i}B^nY^{-i}C^n|X_i=x, Y_i=y, Z_i=z}
\]

which can be implemented by a No-Signaling function and satisfy:
\[
\sum_{i=1}^{n-m} ||P_{XYZQ_{X^{-i}A^nY^{-i}B^nY^{-i}C^n}|XYZ} - P_{XYZX^{-i}A^nY^{-i}B^nY^{-i}C^n}|W|| \leq 13 \sqrt{(n - m) \log \left( \frac{1}{\Pr[W]} \right)}.
\]

Therefore, there exists a coordinate $i_m$ such that
\[
||P_{XYZQ_{X^{-i_m}A^nY^{-i_m}B^nY^{-i_m}C^n|XYZ} - P_{XYZX^{-i_m}A^nY^{-i_m}B^nY^{-i_m}C^n}|W|| \leq 13 \sqrt{\frac{1}{n - m} \log \left( \frac{1}{\Pr[W]} \right)}.
\]

Hence, given input $(X, Y, Z)$ for the game $G$, Alice, Bob and Charlie can use the No-Signaling Strategy for $G^\otimes n$ to obtain a No-Signaling strategy for $G$. The provers play $G$ in the $i_m$ coordinate and answer $A_{i_m}, B_{i_m}, C_{i_m}$ (and ignoring the redundant information).

Thus there is a coordinate $i_m$ on which
\[
\Pr[W_{i_m}|W] \leq \omega^\otimes (G) + 13 \sqrt{\frac{1}{n - m} \log \left( \frac{1}{\Pr[W]} \right)}.
\]

\[\square\]

**Theorem 11.** For all games $G = (X, Y, Z, A, B, C, P_{XYZ}, Q)$ and any positive integer $n$,
\[
\omega^\otimes (G^\otimes n) \leq \left( 1 - \frac{(1 - \omega^\otimes (G))^2}{4000} \right)^n
\]

97
Proof. Fix any strategy for the \( n \)-folded parallel repetition game \( G^\otimes n \) and let \( q_m = \Pr[W_1 \land \ldots \land W_{m}] \). By Proposition 5.5.1 we obtain:

\[
q_{m+1} \leq q_m \cdot \left( \omega^\Diamond (G) + 13 \sqrt{\frac{1}{n-m} \log \left( \frac{1}{q_m} \right)} \right)
\]

We show by induction that \( q_{m+1} \leq \left( \frac{1 + \omega^\Diamond (G)}{2} \right)^{m+1} \) for any \( m + 1 \leq \frac{(1 - \omega^\Diamond (G))(n-m)}{1000} \). If \( q_m \leq \left( \frac{1 + \omega^\Diamond (G)}{2} \right)^{m+1} \), then the claim immediately follows. If \( q_m > \left( \frac{1 + \omega^\Diamond (G)}{2} \right)^{m+1} \), then,

\[
\log \left( \frac{1}{q_m} \right) \leq \log \left( \frac{1 + \omega^\Diamond (G)}{2} \right)^{-(m+1)}.
\]

Since for any \( \alpha, \beta < 1 \), \((1 - \alpha)^{b} \leq 1 - \alpha \beta\), then \((1 - \frac{1}{2})^{1 - \omega^\Diamond (G)} \leq (1 - \frac{1 - \omega^\Diamond (G)}{2})\) and we obtain that

\[
\log \left( \frac{1 + \omega^\Diamond (G)}{2} \right)^{-(m+1)} = \log \left( 1 - \frac{1 - \omega^\Diamond (G)}{2} \right)^{-(m+1)}
\]

\[
\leq \log \left( \frac{1}{2} \right) = (m + 1)(1 - \omega^\Diamond (G))
\]

Thus for any \( m \) satisfying \( m + 1 \leq \frac{(1 - \omega^\Diamond (G))(n-m)}{1000} \),

\[
q_{m+1} \leq q_m \cdot \left( \omega^\Diamond (G) + 13 \sqrt{\frac{1}{n-m} (m + 1)(1 - \omega^\Diamond (G))} \right) \leq q_m \cdot \left( \frac{1 + \omega^\Diamond (G)}{2} \right)
\]

combining the induction hypothesis we obtain \( q_m \cdot \left( \frac{1 + \omega^\Diamond (G)}{2} \right)^{m+1} \). Taking \( m = n \frac{(1 - \omega^\Diamond (G))}{2000} \) we get,

\[
q_m \leq \left( \frac{1 - \omega^\Diamond (G)}{2} \right)^{\frac{n(1 - \omega^\Diamond (G))}{2000}} \leq \left( 1 - \frac{(1 - \omega^\Diamond (G))^2}{4000} \right)^n.
\]

\( \square \)

5.6 Discussion

Our upper bound works only for our definition of No-Signaling (as defined in Section 5.1). Our techniques would not work for the standard version of the generalization of No-Signaling for \( k \) players. Recall that in the standard model, the No Signaling condition
ensures that the answer of each player given its question, is independent of the questions of the other players. More formally, in the standard model, a function

\[ p_1 : X \times Y \times Z \times R \rightarrow A \]

is a No-Signaling function if:

\[ \forall x \in X, a \in A, y, y' \in Y, z, z' \in Z \quad \Pr_R[p_1(x, y, z, R) = a] = \Pr_R[p_1(x, y', z', R) = a], \]

\[ \forall x \in X, a \in A, y \in Y, z, z' \in Z \quad \Pr_R[p_1(x, y, z, R) = a] = \Pr_R[p_1(x, y, z', R) = a], \]

\[ \forall x \in X, a \in A, y, y' \in Y, z \in Z \quad \Pr_R[p_1(x, y, z, R) = a] = \Pr_R[p_1(x, y', z, R) = a]. \]

In Lemma 5.4.4, we showed that given a function that satisfies the conditions in Equations (5.5, 5.6, 5.7), one can obtain a function that is No-Signaling in the non-standard model but is not necessarily No-Signaling in the standard case. To understand why Lemma 5.4.4 does not hold for the standard model, consider the uniform distribution over \( \{0, 1\} \times \{0, 1\} \times \{0, 1\} \). Assume that given the answers of the players on the \( 1, \ldots, m - 1 \) games, the conditional distribution on the \( i \)'s coordinate is

\[ \{(0, 0, 0), (1, 1, 0), (1, 0, 1), (0, 1, 1)\}. \]

Notice that the marginal distribution of the questions of player 1 (that is uniform over \( \{0, 1\}\)) is identical to the marginal distribution over her questions, conditioned on any question of the other players. However, this is not the case when conditioning the marginal distribution of the questions of player 1, on the questions of both player 2 and player 3. Given the questions of any two players, the question of the third player is just the XOR of the other questions. Therefore, the definition used in this chapter is crucially to our analysis. It would be very interesting to find a new technique that would work for the standard definition of No-Signaling.
Chapter 6

Discussion and Open Problems

6.1 Concluding remarks

In this thesis, we have shown the following:

• We proved two technical lemmas:

  − We showed that if the relative entropy of \( P \) with respect to \( Q \) is small, then for every set \( S \), whose measure according to \( P \) is not too small, \( Q(S) \) is not much bigger than \( P(S) \).

  − We showed that, if two distributions are ‘close’ from the point of view of an average vertex, then they are ‘close’ to each other on the entire space.

• We prove parallel repetition theorems for games played on a (biregular) bipartite expander graph whose normalized second singular value is \( \lambda \).

  − We prove that the value of the repeated (general) game is at most

\[
(1 - \epsilon^2)^{\Omega(c(\lambda) \cdot n/s)},
\]

where \( c(\lambda) = \text{poly}(\lambda) \).
6.2 Non Product Distribution Between Coordinates

We are interested in parallel repetition theorems for games where the distribution between the coordinates is not a product distribution.

The parallel repetition of a game $G$ is a game where the verifier generates questions $x = (x_1, \ldots, x_n) \in X^n$, $y = (y_1, \ldots, y_n) \in Y^n$, where each pair $(x_i, y_i) \in X \times Y$ is chosen independently according to the original distribution $P_{XY}$, that is, $P_{X^nY^n} = P_{XY} \times \ldots \times P_{XY}$. We would like to consider the case where the pairs $(x_i, y_i) \in X \times Y$ are not chosen...
independently. Is there a distribution according to which the verifier generates questions $x = (x_1, \ldots, x_n) \in X^n$, $y = (y_1, \ldots, y_n) \in Y^n$ for which a better bound on the value of the game holds? Is there a distribution for which even a strong parallel repetition holds?

### 6.3 Entangled Provers

An entangled (or quantum) two-prover game is played between two players called provers and an additional player called a verifier. The verifier in this model is the same verifier as in the classical model but the provers in this model may use a quantum protocol. We now explain the term 'quantum protocol'. In this model, the two provers share between them an arbitrary entangled quantum state and each prover may measure her part of the state before answering the question addressed to her by the verifier. Before the games begin the provers prepare any quantum state divided into two parts. One part for each prover. When the games begins, each prover receives a question from the verifier and each prover knows only the question addressed to her. The provers may now measure their part of the quantum state and can use this measurement to answer the verifier. The basis according to which the measurement is taken, can depend on the questions that the provers received. This property enables the provers to create nontrivial correlations between their answers. The entangled (or quantum) value of the game is defined to be

$$\max_{a,b} \mathbb{E}_{(x,y)}[V(x, y, a(x), b(y))],$$

where the maximum is taken over all quantum protocols $(a, b)$ and the expectation is taken with respect to the distribution of the questions. There are no parallel repetition theorems for entangled two-prover games. However, recently Kempe and Regev [25] showed that strong parallel repetition does not hold with entangled provers.

It would be very interesting to give an accurate description of the behavior of the value of a repeated quantum game.
6.4 More than two player in other models

In a $k$-provers one-round game there are $k$ provers and a verifier. The verifier selects randomly $(x_1, \ldots, x_k) \in X_1 \times \cdots \times X_k$, a question for each prover, according to some distribution $P_{X_1 \ldots X_k}$ where $X_i$ is the questions set of prover $i$ where $i \in 1, \ldots, k$. Each prover knows only the question addressed to her, prover 1 knows only $x_1$ and prover $i$ knows only $x_i$. The provers cannot communicate during the transaction. The provers send their answers to the verifier, $a_i \in A_i$ where $A_i$ is the answers set of prover $i$. The verifier evaluates an acceptance predicate $V(x_1, \ldots, x_k, a_1, \ldots, a_k)$ and accepts or rejects based on the outcome of the predicate. The acceptance predicate as well as the distribution of the questions are known in advance to the provers. The provers answer the questions according to a strategy.

There are no nontrivial parallel repetition theorems for more than two provers in the classical model. It would be interesting to consider this problem both in the classical model and in other models.

6.5 Unique Games

The unique games conjecture (UGC) by Khot [26] states that for every sufficiently small pair of constants $\epsilon, \delta > 0$, there exists a constant $k(\epsilon, \delta)$ such that given a unique game $G$, it is NP-hard to distinguish between the case that the value of $G$ is at least $1 - \delta$ from the case that the value of $G$ is at most $\epsilon$, where $k$ is the size of the answers set of $G$.

Since Khot suggested the UGC, a massive research was done due to its strong consequences for hardness of approximation. Many connections between computational complexity, algorithms, analysis, and geometry were made in the purpose of understanding this conjecture.

One of the leading strategies towards understanding the UGC was suggested by Feige, Kindler, and ODonnell [19]. The authors showed that a strong parallel repetition theorem
(or even ‘close’ to strong) could be extremely helpful in studying the unique games conjecture. Since a strong parallel repetition theorem is not true in general [33], it is interesting to try to prove it for general subclasses of games.

It could be the case that for a subclass of games or even for a specific interesting game, a strong parallel repetition holds and this could be meaningful in understanding the unique games conjecture.

We note, however, that unique games on expander graphs are easy [6, 29, 28].
Chapter 7

Appendix

7.1 Analysis of the Boundaries in the Proof of Lemma (2.3.2)

In this section, we consider the case that for some \((x_0, y_0)\), \(P(x_0, y_0) = 0\) or \(P(x_0, y_0) = 1\), in the proof of Lemma (2.3.2) (see Footnote 1).

Case 1: There exists \((x_0, y_0)\) for which \(P(x_0, y_0) = 1\); then it must be that \(P(x_0) = 1\) and \(P(y_0) = 1\), thus, Lemma (2.3.2) holds by taking \(c_0 = 0\) and \(c_1 = 1\) or vice-versa.

Case 2: There exists a distribution \(P\) that minimizes

\[
\mathbb{E}_{Y \sim P_Y} D(P_{X|Y} \| Q_{X|Y}) + \mathbb{E}_{X \sim P_X} D(P_{Y|X} \| Q_{Y|X})
\]

(7.1)

(under the constraints), for which there exists \((x_0, y_0)\) such that \(P(x_0, y_0) = 0\) and \(P(x_0) > 0\). Without loss of generality, assume \((x_0, y_0) \in S\). Let \((x_1, y_1) \in S\) be such that \(P(x_1, y_1) > 0\). Denote \(\tau = P(x_1, y_1)\).

By definition,

\[
\mathbb{E}_{X \sim P_X} D(P_{Y|X} \| Q_{Y|X}) = \sum_{(x,y) \in X \times Y} P(x,y) \log \left( \frac{P(x,y)/P(x)}{Q(x,y)/Q(x)} \right)
\]
and similarly,
\[ E_{Y \sim P_Y} D(P_{X|Y} \| Q_{X|Y}) = \sum_{(x,y) \in X \times Y} P(x,y) \log \left( \frac{P(x,y)/P(y)}{Q(x,y)/Q(y)} \right) \]

Note that the derivative of Equation (7.1) in the variable \( P(x_0, y_0) \) at the point \( P(x_0, y_0) = 0 \) is \( -\infty \), while the derivative of Equation (7.1) in the variable \( P(x_1, y_1) \) at the point \( P(x_1, y_1) = \tau \) is finite. Thus, if we move a tiny mass from \( P(x_1, y_1) \) to \( P(x_0, y_0) \), the new distribution would decrease Equation (7.1). This is a contradiction to the assumption that \( P \) minimizes Equation (7.1).

A symmetric argument shows that the case where there exists \((x_0, y_0)\) such that \( P(x_0, y_0) = 0 \) and \( P(y_0) > 0 \) does not minimize Equation (7.1).

**Case 3:** There exists a distribution \( P \) that minimizes Equation (7.1) (under the constraints), such that for all \((x_0, y_0)\) for which \( P(x_0, y_0) = 0 \), we have
\[ P(y_0) = 0, \quad P(x_0) = 0. \]

Consider the set of all the points that satisfy the conditions above and denote this set by \( T \). That is,
\[ T := \{ (x_0, y_0) \mid P(y_0) = 0, \ P(x_0) = 0 \}. \]

The set \( T \) is a set of edges in the graph. Every edge \((x, y)\) in the graph that shares a vertex with an edge in \( T \), that is, either there exists \( y_0 \) such that \((x, y_0) \in T \) or there exists \( x_0 \) such that \((x_0, y) \in T \), must satisfy \( P(x) = 0 \) or \( P(y) = 0 \); and hence, \( P(x, y) = 0 \). Therefore, since the graph is connected, \( T \) is either the set of all edges in the graph (thus \( P \) is not a distribution) or the empty set.
7.2 Proof of Lemma 5.4.1

Lemma 7.2.1 (Raz). Let \( P_{U\ell} = P_{U_1}P_{U_2} \cdots P_{U_\ell} \) and let \( W \) be some event, then the following holds:

\[
\sum_{i=1}^{\ell} \|P_{U_i|W} - P_{U_i}\| \leq \ell \cdot \log \frac{1}{\Pr[W]}
\]

Proof. This lemma appears in [32] and in [24]. We present Holenstein’s [24] proof of the lemma. For the proof of the lemma we’ll need to use some known entropy arguments such as \( D(P||Q) \geq (\|P - Q\|_1)^2 \) where \( P \) and \( Q \) are two distributions over the same sampling space. The proof can be found in [13] Lemma 12.6.1. Another argument we will use is that for a probability distribution \( P_{U\ell} \) which is a product distribution over \( U_1 \ldots U_\ell \) the following holds:

\[
\sum_{i=1}^{\ell} D(P_{U_i||Q_{U_i}}) \leq D(P_{U\ell||Q_{U\ell}})
\]

We can now prove the lemma:

\[
\begin{align*}
\sum_{i=1}^{\ell} (\|P_{U_i|W} - P_{U_i}\|)^2 &\leq \sum_{i=1}^{\ell} D(P_{U_i|W||P_{U_i}}) \leq D(P_{U\ell|W||P_{U\ell}}) \\
&= \sum_{u^\ell \in U^\ell} P_{U\ell|W}(u^\ell) \log \left( \frac{P_{U\ell|W}(u^\ell)}{P_{U\ell}(u^\ell)} \right) \\
&= \sum_{u^\ell \in U^\ell} P_{U\ell|W}(u^\ell) \log \left( \frac{\Pr[W|U^\ell = u^\ell]}{\Pr[W]} \right) \\
&\leq \log \left( \frac{1}{\Pr[W]} \right) + \sum_{u^\ell \in U^\ell} P_{U\ell|W}(u^\ell) \log(\Pr[W|U^\ell = u^\ell]) \\
&\leq \log \left( \frac{1}{\Pr[W]} \right)
\end{align*}
\]
7.3 Proof of Corollary 5.4.2

Proof.

\[
\sum_{i=1}^{\ell} \| P_{T|W_i} - P_{T|W}P_{U_i|T} \| = \sum_{i=1}^{\ell} \sum_{t \in T, u \in U_i} | P_{T|W_i}(t, u) - P_{T|W}(t)P_{U_i|T=t}(u) |
\]

\[
= \sum_{i=1}^{\ell} \sum_{t \in T, u \in U_i} | P_{T|W}(t)P_{U_i|W,T=t}(u) - P_{T|W}(t)P_{U_i|T=t}(u) |
\]

\[
= \sum_{t \in T} P_{T|W}(t) \cdot \sum_{i=1}^{\ell} \| P_{U_i|W,T=t} - P_{U_i|T=t} \|
\]

\[
\leq \sum_{t \in T} P_{T|W}(t) \cdot \sqrt{\ell \cdot \log \left( \frac{1}{\Pr[W|T = t]} \right)}
\]

\[
\leq \sqrt{\ell \cdot \log \left( \sum_{t \in T} P_{T|W}(t) \cdot \frac{1}{\Pr[W|T = t]} \right)}
\]

\[
= \sqrt{\ell \cdot \log \left( \sum_{t \in T} \frac{\Pr(T = t \land W)}{\Pr[W]} \cdot \frac{\Pr[T = t]}{\Pr[W \land T = t]} \right)}
\]

\[
= \sqrt{\ell \cdot \log \left( \frac{1}{\Pr[W]} \right)}
\]

Where the first inequality follows from Lemma 5.4.1 and the second follows from Jensen’s inequality on the concave function \( \sqrt{\log(\cdot)} \) \( \square \)

7.4 Generalizing for \( k > 3 \)

Lemma 7.4.1. Let \( P_{V_1,\ldots,V_k} \) be a probability distribution over \( V_1 \ldots V_k \) and for every \( 1 \leq i \leq k \), let \( P_{\bar{V}_i} \) be some distribution over \( V_i \). If for every \( i \in \{1, \ldots, k\} \), \( \| P_{V_i} - P_{\bar{V}_i} \| \leq \epsilon_i \) then, there exists a distribution \( Q_{V_1,\ldots,V_k} \) satisfying the following:

\[
\| Q_{V_1,\ldots,V_k} - P_{V_1,\ldots,V_k} \| \leq 2 \sum_{i=1}^{k} \epsilon_i + \min_i \epsilon_i
\]

and for every \( i \) \( \| Q_{\bar{V}_i} - P_{\bar{V}_i} \| = 0 \)
Lemma 7.4.2. Let \( P_{U_1, \ldots, U_k, V_1, \ldots, V_k} \) be a probability distribution over \( U_1, \ldots, U_k, V_1, \ldots, V_k \) and let \( P'_{U_1, \ldots, U_k} \) be some distribution over \( U_1, \ldots, U_k \). If for every \( i \),

\[
\|P'_{U_1, \ldots, U_k} P_{V_i | U_i} - P_{U_1, \ldots, U_k, V_i}\| \leq \epsilon_i
\]

then, there exists a conditional distribution \( Q_{V_1, \ldots, V_k | X=x, Y=y, Z=z} \) satisfying the following:

\[
\|Q_{V_1, \ldots, V_k | U_1=u_1, \ldots, U_k=u_k} - P_{V_1, \ldots, V_k | U_1=u_1, \ldots, U_k=u_k}\| \leq \min_i \{\epsilon_i\} \quad (7.2)
\]

\[
\forall i \in \{1, \ldots, k\}, \quad \|Q_{V_i | U_1=u_1, \ldots, U_k=u_k} - Q_{V_i | U_i=u_i}\| = 0 \quad (7.3)
\]

\[
(7.4)
\]

Theorem 12. For every \( k \geq 3 \), every positive integer \( n \), all games

\[
G = (X, Y, Z, A, B, C, P_{X,Y,Z}, Q)
\]

, played in the MIP\( ^\diamond (k, 1) \) model satisfy:

\[
\omega^\diamond (G^{\otimes n}) \leq \left(1 - \frac{1 - \omega^\diamond (G)^2}{100(1 + 4k)^2}\right)^n
\]

Proof. Fix any strategy for the \( n \)-folded parallel repetition game \( G^{\otimes n} \) and let

\[
q_m = \Pr[W_1 \land \ldots \land W_m].
\]

By Proposition 5.5.1 we obtain:

\[
q_{m+1} \leq q_m \cdot \left(\omega^\diamond (G) + 13 \sqrt{\frac{1}{n-m} \log \left(\frac{1}{q_m}\right)}\right)
\]

We show by induction that \( q_{m+1} \leq \left(\frac{1+\omega^\diamond (G)}{2}\right)^{m+1} \) for any \( m + 1 \leq \frac{(1-\omega^\diamond (G))(n-m)}{1000} \). If \( q_m \leq \left(\frac{1+\omega^\diamond (G)}{2}\right)^{m+1} \) then the claim immediately follows. If \( q_m > \left(\frac{1+\omega^\diamond (G)}{2}\right)^{m+1} \) then,

\[
\log \left(\frac{1}{q_m}\right) \leq \log \left(\frac{1 + \omega^\diamond (G)}{2}\right)^{-(m+1)}.
\]
7.4. GENERALIZING FOR $K > 3$

Since for any $\alpha, \beta < 1$, $(1 - \alpha)^b \leq 1 - \alpha \beta$, then $(1 - \frac{1}{2})^{1-\omega^{\diamond}(G)} \leq \left(1 - \frac{1-\omega^{\diamond}(G)}{2}\right)$ and we obtain that

$$\log\left(\frac{1 + \omega^{\diamond}(G)}{2}\right)^{(m+1)} = \log\left(1 - \frac{1 - \omega^{\diamond}(G)}{2}\right)^{(m+1)} \leq \log\left(\frac{1}{2}\right)^{(m+1)} = (m + 1)(1 - \omega^{\diamond}(G))$$

Thus for any $m$ satisfying $m + 1 \leq \frac{(1-\omega^{\diamond}(G))(n-m)}{1000}$,

$$q_{m+1} \leq q_m \cdot \left(\omega^{\diamond}(G) + 13\sqrt{\frac{1}{n-m}(m + 1)(1 - \omega^{\diamond}(G))}\right) \leq q_m \cdot \left(\frac{1 + \omega^{\diamond}(G)}{2}\right)$$

combining the induction hypothesis we obtain $q_m \cdot \left(\frac{1+\omega^{\diamond}(G)}{2}\right) \leq \left(\frac{1+\omega^{\diamond}(G)}{2}\right)^{m+1}$. Taking $m = n\frac{(1-\omega^{\diamond}(G))}{2000}$ we get,

$$q_m \leq \left(1 - \frac{1 - \omega^{\diamond}(G)}{2}\right)^{n\frac{(1-\omega^{\diamond}(G))}{2000}} \leq \left(1 - \frac{(1 - \omega^{\diamond}(G))^2}{4000}\right)^n$$

□
Bibliography


