Electronic Colloquium on Computational Complexity, Report No. 142 (2010)

A Strong Parallel Repetition Theorem for Projection Games on Expanders

Ran Raz* Ricky Rosen †

Abstract

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 [24, 17]. For *Projection Games*, the bound on the value of the game repeated n times in parallel was improved to $(1-\epsilon^2)^{\Omega(n)}$ [23] and this bound was shown to be tight [25].

In this paper 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 show that if λ is the (normalized) spectral gap of the underlying graph, the value of the repeated game is at most

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

where $c(\lambda) = \text{poly}(\lambda)$; and if in addition the game is a projection game, we obtain a bound of

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

where $c(\lambda) = \text{poly}(\lambda)$, that is, a *strong parallel repetition theorem* (when λ is constant).

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

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 $V : X \times Y \times A \times B \to \{0,1\}$. 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 $V(x,y,a,b) = 1$.

*Faculty of mathematics and computer science, Weizmann Institute. Research supported by the Israel Science Foundation (ISF).

†Department of Computer Science, Tel-Aviv University.

ISSN 1433-8092

The provers answer the questions according to a pair of functions $a : X \to A$, $b : Y \to B$. 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.

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 $f_{x,y} : B \to A$, 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*.

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.

The parallel repetition theorem [24] states that for any two-prover game G, with value $\leq 1 - \epsilon$ (for any $0 < \epsilon \leq 1/2$), the value of the game $G^{\otimes n}$ is at most

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

where s is the answers' length of the original game, and c is a universal constant. The constant c implicit in [24] is $c = 32$. An example by Feige and Verbitsky [15] shows that the dependency on s in Equation (1) is necessary.

A beautiful recent work by Holenstein [17] simplified the proof of [24] and obtained an improved constant of $c = 3$. An intriguing followup work by Rao [23] gave for the important special case of projection games, an improved bound of

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

Several researchers asked wether or not these bounds could be improved to $(1 - \epsilon)^{\Omega(n/s)}$, for general two-prover games, or at least for interesting special cases, such as, projection games, unique games, or xor games (see for example [13, 27]); this question is usually referred to as the *strong parallel repetition problem*. However, a recent analysis shows that the, so called, *odd cycle game* (first studied in [13, 11]) is a counterexample to strong parallel repetition [25]. 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 $\geq (1 - \epsilon^2)^{O(n)}$ [25] (see also [9]). 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 (2) is tight.

The parallel repetition theorem was used to prove a large number of hardness of approximation results [8], such as, Håstad's optimal results on the hardness of approximation of 3-SAT and 3-LIN [16], and Feige's optimal results on the hardness of approximation of Set-Cover [12]. It also has applications in quantum information theory [11]; in understanding foams and tiling the space $\mathbb{R}^n$ [13, 20]; and in communication complexity [22, 7]. For farther discussion see [26].

1.3 Our Results

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, λ 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 λ is the (normalized) spectral gap of the graph, the value of the repeated game is at most

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

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

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

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

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

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 note also that in many applications, and in particular in most applications for hardness of approximation, the underlying graph is a (biregular) bipartite expander graph.

1.4 Related Works

Strong Parallel Repetition Theorem for Free Projection Games:

For games where the distribution P on $X \times Y$ is a product distribution, bounds of $(1-\epsilon^2)^{\Omega(n/s)}$ (for general games) and $(1-\epsilon)^{\Omega(n)}$ (for projection games), were previously obtained [10]. 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) [3] (using [14]). More precisely, they obtained a bound of $(1-\frac{\epsilon}{\log(1/\epsilon)})^{\Omega(\lambda \cdot n)}$.

Safra and Schwartz [27] 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.

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 [18]. 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 [2, 1]. It turned out that a strong parallel repetition theorem (or even 'close' to strong) could be extremely helpful in studding the unique games conjecture. Since, a strong parallel repetition theorem is not true in general [25], it is interesting to try to prove it for general subclasses of games.

We note however that unique games on expander graphs are easy [3, 21, 19]. 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.

1.5 Techniques

Our proof goes along the general lines of the original proof of the parallel repetition theorem [24]. 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 ϵ, that we cannot afford.

In previous results [24, 17, 23], 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 ℓ_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 ℓ_1 norm instead of using directly the relative entropy. A recent counterexample shows that in the general case, this loss is necessary [25].

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 ℓ_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 ϵ 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 [24]. The proof for the case of projection games requires the use of Rao's ideas [23] together with additional new techniques. In particular, for the case of projection games we use once again the expansion properties of the underlying graph.

2 Preliminaries

2.1 Notations

2.1.1 General Notations

We denote an n-dimensional vector by a superscript n, e.g., $\phi^n = (\phi_1, \ldots, \phi_n)$ where ϕ_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 $\{1, \ldots, n\}$.

2.1.2 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 (4.1), to think of $\Pr(Z)$ as a random variable.

2.1.3 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 $\mathrm{P}_{X|U}$ 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 $\mathrm{P}_{X|Z}$ 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 $\mathrm{P}_{XY|U}$ to denote the distribution of X, Y conditioned on the event U, i.e., $\mathrm{P}_{XY|U}(x, y) = \Pr(X = x, Y = y|U)$. 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.

2.1.4 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 $\mathrm{G}(\mathrm{P}_{XY}, V)$. 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.

2.1.5 The Game G Repeated n Times

For the game G repeated n times in parallel, $\mathrm{G}^{\otimes n} = G(\mathrm{P}_{X^nY^n}, V^{\otimes n})$, 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 $\mathrm{G}^{\otimes n}$, 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, $Y^{n-k} = (Y_1, \ldots, Y_{n-k})$.

2.1.6 The Event W_i

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

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

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 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 $f_{xy}(b) = a$.*

2.2 Entropy and Relative Entropy

Definition 2.2 (Entropy)**.** *For a probability distribution ϕ over a sample space Ω we define the entropy of ϕ to be* $\mathrm{H}(\phi) = -\sum_{x\in\Omega} \phi(x) \log \phi(x) = -\mathbb{E}_{x\sim\phi} \log \phi(x) = \mathbb{E}_{x\sim\phi} \log\left(\frac{1}{\phi(x)}\right)$

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

Fact 2.3. *For every distribution ϕ over Ω,* $\mathrm{H}(\phi) \leq \log(|\mathrm{supp}(\phi)|)$ *where*

$$\mathrm{supp}(\phi) = \{x \in \Omega | \phi(x) > 0\}$$

Definition 2.4 (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* Ω*. The relative entropy of* P *with respect to* Q *is:*

$$\mathrm{D}(\mathrm{P}\|\mathrm{Q}) = \sum_{x \in \Omega} \mathrm{P}(x) \log \frac{\mathrm{P}(x)}{\mathrm{Q}(x)}$$

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* ∞*.*

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 apriority 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 2.5. *Let* $\Phi^n = \Phi_1 \times \Phi_2 \times \cdots \times \Phi_n$ *and let* μ^n *be any distribution over the same sample space (not necessarily a product distribution) then* $\sum_{i=1}^{n} \mathrm{D}(\mu_i \| \Phi_i) \leq \mathrm{D}(\mu^n \| \Phi^n)$ *thus* $\mathbb{E}_{i \in [n]} \mathrm{D}(\mu_i \| \Phi_i) = \frac{1}{n} \sum_{i \in [n]} \mathrm{D}(\mu_i \| \Phi_i) \leq \frac{\mathrm{D}(\mu^n \| \Phi^n)}{n}$

2.3 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 our paper 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.

2.3.1 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 Σ 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 Σ 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 σ 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 σ, respectively.

In any singular value decomposition $M = U\Sigma V^*$ the diagonal entries of Σ 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.

2.3.2 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 σ_0. We refer to it as the *second singular value*. We say that λ is the *spectral gap* of G_{XY}.

2.3.3 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 $1-\lambda$.

2.4 Technical Lemma

The following lemma appeared first at [10]

Lemma 2.6. *For every $0 \leq p, q \leq 1$ define binary distributions $\mathrm{P} = (p, 1-p)$ and $\mathrm{Q} = (q, 1-q)$, over $\{0,1\}$, if $\mathrm{D}(\mathrm{P}\|\mathrm{Q}) \leq \delta$ and $p < \delta$ 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 $q > p$, $\mathrm{D}(\mathrm{P}\|\mathrm{Q})$ is decreasing in p and increasing in q,

$$\begin{aligned} \mathrm{D}(\mathrm{P}\|\mathrm{Q}) &= p\log\frac{p}{q} + (1-p)\log\frac{1-p}{1-q} \\ &> \delta\log(\frac{\delta}{4\delta}) + (1-\delta)\log\frac{1-\delta}{1-4\delta} \\ &= -2\delta + (1-\delta)\log\left(1 + \frac{3\delta}{1-4\delta}\right) \end{aligned} \tag{3}$$

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

$$(3) \geq -2\delta + (1-\delta) > \delta$$

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,

$$(3) \quad \geq \quad -2\delta + (1-\delta)\frac{3\delta}{1-4\delta} \geq -2\delta + 3\delta = \delta$$

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 2.7. *For every probability distributions* P, Q *over the same sample space* Ω *and for every* $T \subseteq \Omega$, *if* $\mathrm{D}(\mathrm{P}\|\mathrm{Q}) \leq \delta$ *and* $\mathrm{P}(T) \leq \delta$ *then* $\mathrm{Q}(T) \leq 4\delta$

Proof. Denote $p = \mathrm{P}(T)$ and $q = \mathrm{Q}(T)$ and let $\mathrm{P}' = (p, 1-p)$, $\mathrm{Q}' = (q, 1-q)$. By the data processing inequality for mutual information $\mathrm{D}(\mathrm{P}\|\mathrm{Q}) \geq \mathrm{D}(\mathrm{P}'\|\mathrm{Q}')$ and the corollary follows. □

3 The Main Lemma

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

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

For simplicity, we write $\mathrm{P}(x)$ rather than $\mathrm{P}_X(x)$. Similarly, we use $\mathrm{Q}_X, \mathrm{P}_Y, \mathrm{Q}_Y$, for the marginal distribution of Q on X, the marginal distribution of P on Y and the marginal distribution of Q on Y respectively. For $y \in Y$, the distribution $\mathrm{P}_{X|y}$ is the marginal distribution of P on X conditioned on $Y = y$, i.e.,

$$\mathrm{P}_{X|y}(x) = \mathrm{P}(x|y) = \mathrm{P}(x, y)/\mathrm{P}(y).$$

We use $\mathrm{Q}_{X|y}$ for the marginal distribution of Q on X conditioned on $Y = y$. For $x \in X$ we use $\mathrm{P}_{Y|x}$ and $\mathrm{Q}_{Y|x}$ 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 $\mathrm{P}_{X|Y}$ 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 (3.1)

$$\mathbb{E}_{Y \sim \mathrm{P}_Y} \mathrm{D}(\mathrm{P}_{X|Y} \| \mathrm{Q}_{X|Y})$$

this is the expected value of the relative entropy of $\mathrm{P}_{X|Y}$ with respect to $\mathrm{Q}_{X|Y}$ where Y is distributed according to P_Y. Similarly, $\mathrm{Q}_{X|Y}$ for the marginal distribution of Q on X conditioned on Y where Y is a random variable distributed over the set Y, $\mathrm{P}_{Y|X}$ for the marginal distribution of P on Y conditioned on X where X is a random variable distributed over the set X and $\mathrm{Q}_{Y|X}$ for the marginal distribution of Q on Y conditioned on X where X is a random variable distributed over the set X.

Lemma 3.1. *Let X, Y be two sets and let d_X, d_Y be two integers. Let* G *be a $(X, Y, d_X, d_Y, 1-\lambda)-$bipartite expander graph with second singular value $1-\lambda$ (for $\lambda > 0$). 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:*

$$\mathbb{E}_{Y\sim \mathrm{P}_Y} \mathrm{D}(\mathrm{P}_{X|Y} \| \mathrm{Q}_{X|Y}) \leq \epsilon$$

and

$$\mathbb{E}_{X\sim \mathrm{P}_X} \mathrm{D}(\mathrm{P}_{Y|X} \| \mathrm{Q}_{Y|X}) \leq \epsilon$$

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

$$\mathrm{Q}(S) < 10^6 \cdot \alpha \cdot \frac{1}{\lambda} \cdot \log \frac{2}{\lambda}$$

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

$$\alpha' := \mathrm{Q}(S).$$

Let P be a distribution over E_{G}. We will show that if $\mathrm{P}(S) = \alpha$, and

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

then

$$\mathbb{E}_{Y\sim \mathrm{P}_Y} \mathrm{D}(\mathrm{P}_{X|Y} \| \mathrm{Q}_{X|Y}) + \mathbb{E}_{X\sim \mathrm{P}_X} \mathrm{D}(\mathrm{P}_{Y|X} \| \mathrm{Q}_{Y|X}) > 2\epsilon.$$

To do that we first find a distribution P that satisfies the conditions above, that is $\mathrm{P}(S) = \alpha$, and minimizes

$$\mathbb{E}_{Y\sim \mathrm{P}_Y} \mathrm{D}(\mathrm{P}_{X|Y} \| \mathrm{Q}_{X|Y}) + \mathbb{E}_{X\sim \mathrm{P}_X} \mathrm{D}(\mathrm{P}_{Y|X} \| \mathrm{Q}_{Y|X}).$$

This is done in Lemma (3.2). We then study this distribution and obtain bounds on its values, i.e., bounds on $\mathrm{P}(x, y)$ for all x, y, (this is done in Lemma (3.3)). Later we show that this distribution is close to the uniform distribution over the edges of G (Lemma (3.4)). For this distribution P, we obtain a contradiction.

Lemma 3.2. *Let* P *be a distribution over E_{G} that minimizes*

$$\mathbb{E}_{Y\sim \mathrm{P}_Y} \mathrm{D}(\mathrm{P}_{X|Y} \| \mathrm{Q}_{X|Y}) + \mathbb{E}_{X\sim \mathrm{P}_X} \mathrm{D}(\mathrm{P}_{Y|X} \| \mathrm{Q}_{Y|X})$$

under the constraint $\mathrm{P}(S) = \alpha$. Then there exist constants $c_0, c_1 > 0$ such that for every $(x, y) \in E_{\mathrm{G}}$:

$$\mathrm{P}(x,y) = \begin{cases} c_1 \sqrt{\mathrm{P}(x)} \cdot \sqrt{\mathrm{P}(y)} & (x,y) \in S; \\ c_0 \sqrt{\mathrm{P}(x)} \cdot \sqrt{\mathrm{P}(y)} & (x,y) \notin S \ . \end{cases}$$

Proof. Since Q is a uniform distribution over a $(d_X, d_Y)-$bipartite graph, for all $(x, y) \in E_{\mathrm{G}} \subseteq X \times Y$, $\mathrm{Q}(y|x) = 1/d_X$ and $\mathrm{Q}(x|y) = 1/d_Y$. By definition,

$$\begin{aligned} \mathbb{E}_{X\sim \mathrm{P}_X} \mathrm{D}(\mathrm{P}_{Y|X} \| \mathrm{Q}_{Y|X}) &= \sum_{(x,y)\in X\times Y} \mathrm{P}(x,y) \log \left(\frac{\mathrm{P}(x,y)/\mathrm{P}(x)}{\mathrm{Q}(x,y)/\mathrm{Q}(x)} \right) \\ &= \sum_{(x,y)\in X\times Y} \mathrm{P}(x,y) \log \mathrm{P}(x,y) - \sum_{x\in X} \mathrm{P}(x) \log \mathrm{P}(x) - \log(1/d_X) \end{aligned}$$

and similarly,

$$\begin{aligned}\mathbb{E}_{Y\sim \mathrm{P}_Y}\mathrm{D}(\mathrm{P}_{X|Y}\|\mathrm{Q}_{X|Y}) &= \sum_{(x,y)\in X\times Y}\mathrm{P}(x,y)\log\left(\frac{\mathrm{P}(x,y)/\mathrm{P}(y)}{\mathrm{Q}(x,y)/\mathrm{Q}(y)}\right)\\ &= \sum_{(x,y)\in X\times Y}\mathrm{P}(x,y)\log\mathrm{P}(x,y) - \sum_{y\in Y}\mathrm{P}(y)\log\mathrm{P}(y) - \log(1/d_Y)\end{aligned}$$

Thus,

$$\begin{aligned}&\mathbb{E}_{X\sim \mathrm{P}_X}\mathrm{D}(\mathrm{P}_{Y|X}\|\mathrm{Q}_{Y|X}) + \mathbb{E}_{Y\sim \mathrm{P}_Y}\mathrm{D}(\mathrm{P}_{X|Y}\|\mathrm{Q}_{X|Y}) =\\ &2\sum_{(x,y)\in X\times Y}\mathrm{P}(x,y)\log\mathrm{P}(x,y) - \sum_{x\in X}\mathrm{P}(x)\log\mathrm{P}(x) - \sum_{y\in Y}\mathrm{P}(y)\log\mathrm{P}(y) - \log(1/d_X) - \log(1/d_Y)\end{aligned} \tag{4}$$

We will minimize Equation (4) under the constraints:

$$\begin{aligned}\sum_{(x,y)\in S}\mathrm{P}(x,y) &= \alpha\\ \sum_{(x,y)\notin S}\mathrm{P}(x,y) &= 1-\alpha\end{aligned}$$

where $1 \geq \mathrm{P}(x,y) \geq 0$. We will use Lagrange multipliers to find the minimum[1] of Equation (4). We define the Lagrange function by

$$\begin{aligned}\Lambda = &2\sum_{(x,y)\in X\times Y}\mathrm{P}(x,y)\log\mathrm{P}(x,y) - \sum_{x\in X}\mathrm{P}(x)\log\mathrm{P}(x) - \sum_{y\in Y}\mathrm{P}(y)\log\mathrm{P}(y) - \log(1/d_X) - \log(1/d_Y)\\ &- \lambda_1\left(\sum_{(x,y)\in S}\mathrm{P}(x,y) - \alpha\right) - \lambda_2\left(\sum_{(x,y)\notin S}\mathrm{P}(x,y) - 1 + \alpha\right)\end{aligned}$$

For every $(x,y) \in S$:

$$\frac{\partial\Lambda}{\partial\mathrm{P}(x,y)} = 2\log\mathrm{P}(x,y) - \log\mathrm{P}(x) - \log\mathrm{P}(y) - \lambda_1 = 0$$

For every $(x,y) \notin S$:

$$\frac{\partial\Lambda}{\partial\mathrm{P}(x,y)} = 2\log\mathrm{P}(x,y) - \log\mathrm{P}(x) - \log\mathrm{P}(y) - \lambda_2 = 0$$

Thus for every $(x,y) \in S$:

$$\log\left(\frac{\mathrm{P}(x,y)^2}{\mathrm{P}(x)\cdot\mathrm{P}(y)}\right) = \lambda_1$$

[1]We will use Lagrange multipliers to find the minimum for $1 > \mathrm{P}(x,y) > 0$. In Appendix A we deal with $\mathrm{P}(x,y) = 0$ and $\mathrm{P}(x,y) = 1$.

For every $(x,y) \notin S$:

$$\log\left(\frac{\mathrm{P}(x,y)^2}{\mathrm{P}(x)\cdot\mathrm{P}(y)}\right) = \lambda_2$$

Fixing $c_0 := 2^{(1/2)\cdot\lambda_2}$ and $c_1 := 2^{(1/2)\cdot\lambda_1}$ we obtain:

$$\mathrm{P}(x,y) = \begin{cases} c_1\sqrt{\mathrm{P}(x)}\cdot\sqrt{\mathrm{P}(y)} & (x,y)\in S; \\ c_0\sqrt{\mathrm{P}(x)}\cdot\sqrt{\mathrm{P}(y)} & (x,y)\notin S\ . \end{cases}$$

□

For the rest of the proof, we fix P to be the distribution as in Lemma (3.2), that is, P is the distribution over E_{G} that minimizes

$$\mathbb{E}_{Y\sim\mathrm{P}_Y}\mathrm{D}(\mathrm{P}_{X|Y}\|\mathrm{Q}_{X|Y}) + \mathbb{E}_{X\sim\mathrm{P}_X}\mathrm{D}(\mathrm{P}_{Y|X}\|\mathrm{Q}_{Y|X})$$

under the constraint $\mathrm{P}(S) = \alpha$, and let c_0, c_1 be the constants such that for every $(x,y) \in E_{\mathrm{G}}$:

$$\mathrm{P}(x,y) = \begin{cases} c_1\sqrt{\mathrm{P}(x)}\cdot\sqrt{\mathrm{P}(y)} & (x,y)\in S; \\ c_0\sqrt{\mathrm{P}(x)}\cdot\sqrt{\mathrm{P}(y)} & (x,y)\notin S\ . \end{cases}$$

For the entire proof we denote by δ the smallest non-negative real number for which

$$\sum_{(x,y)\in S}\mathrm{P}(x) - \mathrm{Q}(x) \le \delta\cdot d_X$$
$$\sum_{(x,y)\in S}\mathrm{P}(y) - \mathrm{Q}(y) \le \delta\cdot d_Y$$

Lemma 3.3. *For every $\delta \ge 0$, if the following hold:*

$$\sum_{(x,y)\in S}\mathrm{P}(x) - \mathrm{Q}(x) \le \delta\cdot d_X$$
$$\sum_{(x,y)\in S}\mathrm{P}(y) - \mathrm{Q}(y) \le \delta\cdot d_Y$$

then[2]

$$\frac{1-\alpha}{\sqrt{d_X\cdot d_Y}} \le c_0 \le \frac{1+5\alpha\log\left((\alpha'+\delta)/\alpha\right)}{\sqrt{d_X\cdot d_Y}}$$

and

$$\frac{\alpha/(\alpha'+\delta)}{\sqrt{d_X\cdot d_Y}} \le c_1 \le \frac{8}{\sqrt{d_X\cdot d_Y}}$$

(where $\alpha' := \mathrm{Q}(S)$).

[2]We will only use the upper bound on c_0.

Proof. First we will bound c_0 from below.

$$(1-\alpha)^2 = \left(\sum_{(x,y)\notin S} \mathrm{P}(x,y)\right)^2 = c_0^2 \cdot \left(\sum_{(x,y)\notin S} \sqrt{\mathrm{P}(x)}\cdot\sqrt{\mathrm{P}(y)}\right)^2 \le c_0^2 \cdot \left(\sum_{(x,y)\notin S} \mathrm{P}(x)\cdot \sum_{(x,y)\notin S} \mathrm{P}(y)\right) \tag{5}$$

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 (5) by $c_0^2 \cdot (d_X \cdot d_Y)$. Therefore,

$$c_0 \ge \frac{1-\alpha}{\sqrt{d_X \cdot d_Y}} \tag{6}$$

Next we will bound c_1 from below.

$$\alpha^2 = \left(\sum_{(x,y)\in S} \mathrm{P}(x,y)\right)^2 = c_1^2 \cdot \left(\sum_{(x,y)\in S} \sqrt{\mathrm{P}(x)}\cdot\sqrt{\mathrm{P}(y)}\right)^2 \le c_1^2 \cdot \left(\sum_{(x,y)\in S} \mathrm{P}(x)\cdot \sum_{(x,y)\in S} \mathrm{P}(y)\right) \tag{7}$$

Since

$$\sum_{(x,y)\in S} \mathrm{P}(x) = \sum_{(x,y)\in S} \mathrm{Q}(x) + \sum_{(x,y)\in S} \mathrm{P}(x) - \mathrm{Q}(x) \le d_X(\alpha' + \delta)$$
$$\sum_{(x,y)\in S} \mathrm{P}(y) = \sum_{(x,y)\in S} \mathrm{Q}(y) + \sum_{(x,y)\in S} \mathrm{P}(y) - \mathrm{Q}(y) \le d_Y(\alpha' + \delta)$$

we obtain:

$$\alpha^2 \le c_1^2 \cdot d_X d_Y (\alpha' + \delta)^2.$$

Thus,

$$c_1 \ge \frac{\alpha/(\alpha'+\delta)}{\sqrt{d_X \cdot d_Y}} \tag{8}$$

We now prove upper bounds on c_0, c_1. Using

$$\mathbb{E}_{Y\sim \mathrm{P}_Y} \mathrm{D}(\mathrm{P}_{X|Y} \| \mathrm{Q}_{X|Y}) + \mathbb{E}_{X\sim \mathrm{P}_X} \mathrm{D}(\mathrm{P}_{Y|X} \| \mathrm{Q}_{Y|X}) \le 2\epsilon$$

we obtain:

$$\begin{aligned}
2\epsilon &\ge \mathbb{E}_{Y\sim \mathrm{P}_Y} \mathrm{D}(\mathrm{P}_{X|Y} \| \mathrm{Q}_{X|Y}) + \mathbb{E}_{X\sim \mathrm{P}_X} \mathrm{D}(\mathrm{P}_{Y|X} \| \mathrm{Q}_{Y|X}) \\
&= \sum_{(x,y)\in X\times Y} \mathrm{P}(x,y) \log\left(\frac{\mathrm{P}(x,y)/\mathrm{P}(x)}{\mathrm{Q}(x,y)/\mathrm{Q}(x)}\right) + \sum_{(x,y)\in X\times Y} \mathrm{P}(x,y) \log\left(\frac{\mathrm{P}(x,y)/\mathrm{P}(y)}{\mathrm{Q}(x,y)/\mathrm{Q}(y)}\right) \\
&= \sum_{(x,y)\in S} \mathrm{P}(x,y) \log\left(\frac{\mathrm{P}(x,y)/\mathrm{P}(x)}{\mathrm{Q}(x,y)/\mathrm{Q}(x)}\right) + \sum_{(x,y)\notin S} \mathrm{P}(x,y) \log\left(\frac{\mathrm{P}(x,y)/\mathrm{P}(x)}{\mathrm{Q}(x,y)/\mathrm{Q}(x)}\right) \\
&+ \sum_{(x,y)\in S} \mathrm{P}(x,y) \log\left(\frac{\mathrm{P}(x,y)/\mathrm{P}(y)}{\mathrm{Q}(x,y)/\mathrm{Q}(y)}\right) + \sum_{(x,y)\notin S} \mathrm{P}(x,y) \log\left(\frac{\mathrm{P}(x,y)/\mathrm{P}(y)}{\mathrm{Q}(x,y)/\mathrm{Q}(y)}\right)
\end{aligned} \tag{9}$$

Since Q is a uniform distribution over a $(d_X, d_Y)-$bipartite graph and P is as given in Lemma (3.2), Equation (9)=

$$\begin{aligned}
&\sum_{(x,y)\in S} \mathrm{P}(x,y)\log\left(\frac{c_1\sqrt{\mathrm{P}(y)}/\sqrt{\mathrm{P}(x)}}{1/d_X}\right) + \sum_{(x,y)\notin S} \mathrm{P}(x,y)\log\left(\frac{c_0\sqrt{\mathrm{P}(y)}/\sqrt{\mathrm{P}(x)}}{1/d_X}\right)\\
&+ \sum_{(x,y)\in S} \mathrm{P}(x,y)\log\left(\frac{c_1\sqrt{\mathrm{P}(x)}/\sqrt{\mathrm{P}(y)}}{1/d_Y}\right) + \sum_{(x,y)\notin S} \mathrm{P}(x,y)\log\left(\frac{c_0\sqrt{\mathrm{P}(x)}/\sqrt{\mathrm{P}(y)}}{1/d_Y}\right)\\
&= \sum_{(x,y)\in S} \mathrm{P}(x,y)\log\left(\frac{c_1\sqrt{\mathrm{P}(x)}/\sqrt{\mathrm{P}(y)}}{1/d_Y}\cdot\frac{c_1\sqrt{\mathrm{P}(y)}/\sqrt{\mathrm{P}(x)}}{1/d_X}\right)\\
&+ \sum_{(x,y)\notin S} \mathrm{P}(x,y)\log\left(\frac{c_0\sqrt{\mathrm{P}(x)}/\sqrt{\mathrm{P}(y)}}{1/d_Y}\cdot\frac{c_0\sqrt{\mathrm{P}(y)}/\sqrt{\mathrm{P}(x)}}{1/d_X}\right)\\
&= \sum_{(x,y)\in S} \mathrm{P}(x,y)\log\left(\frac{c_1^2}{1/d_X\cdot 1/d_Y}\right) + \sum_{(x,y)\notin S} \mathrm{P}(x,y)\log\left(\frac{c_0^2}{1/d_X\cdot 1/d_Y}\right)\\
&= \alpha\log\left(c_1^2 d_X d_Y\right) + (1-\alpha)\log\left(c_0^2 d_X d_Y\right).
\end{aligned}$$

Where the last equality follows from $\sum_{(x,y)\in S}\mathrm{P}(x,y)=\alpha$ and $\sum_{(x,y)\notin S}\mathrm{P}(x,y)=1-\alpha$.

Using Equation (6) we obtain:

$$2\epsilon \geq \alpha\log\left(c_1^2\cdot d_X\cdot d_Y\right) + (1-\alpha)\log\left((1-\alpha)^2\right) \geq \alpha\log\left(c_1^2\cdot d_X\cdot d_Y\right) - 3\alpha$$

Thus,

$$\log\left(c_1^2\cdot d_X\cdot d_Y\right) \leq 2\epsilon/\alpha + 3 \leq 5$$

Therefore,

$$c_1 \leq 8/\sqrt{d_X\cdot d_Y}$$

Using Equation (8) we obtain:

$$2\epsilon \geq \alpha\log\left(\alpha^2/(\alpha'+\delta)^2\right) + (1-\alpha)\log\left(c_0^2\cdot d_X\cdot d_Y\right)$$

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

$$\begin{aligned}
\log\left(c_0\cdot\sqrt{d_X\cdot d_Y}\right) &\leq 2\epsilon + 2\alpha\log\left((\alpha'+\delta)/\alpha\right) \leq \log\left(1+4\epsilon+4\alpha\log\left((\alpha'+\delta)/\alpha\right)\right)\\
&\leq \log\left(1+5\alpha\log\left((\alpha'+\delta)/\alpha\right)\right)
\end{aligned}$$

Thus,

$$c_0 \leq \left(1+5\alpha\log\left((\alpha'+\delta)/\alpha\right)\right)/\sqrt{d_X\cdot d_Y}$$

□

We now prove that P is close to the uniform distribution over E_{G}.

Lemma 3.4. *The following hold:*

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

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

$$M_{x,y} = \begin{cases} 1/\sqrt{d_X \cdot d_Y} & \text{if } (x,y) \in E_{\mathrm{G}}; \\ 0 & \text{otherwise.} \end{cases}$$

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

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

Recall that

$$\begin{array}{ll} c_1 = \mathrm{P}(x,y)/\sqrt{\mathrm{P}(x)\mathrm{P}(y)} & \text{for } (x,y) \in S \text{ and } (x,y) \in E_{\mathrm{G}}; \\ c_0 = \mathrm{P}(x,y)/\sqrt{\mathrm{P}(x)\mathrm{P}(y)} & \text{for } (x,y) \notin S \text{ and } (x,y) \in E_{\mathrm{G}}. \end{array} \tag{10}$$

Denote by $\overrightarrow{\sqrt{\mathrm{P}(X)}}$ and $\overrightarrow{\sqrt{\mathrm{P}(Y)}}$ the vectors $\left(\sqrt{\mathrm{P}(x_1)}, \ldots, \sqrt{\mathrm{P}(x_{|X|})}\right)$ and $\left(\sqrt{\mathrm{P}(y_1)}, \ldots, \sqrt{\mathrm{P}(y_{|Y|})}\right)$ respectively.

Claim 3.5. $\overrightarrow{\sqrt{\mathrm{P}(X)}} \cdot \mathrm{M}' = \overrightarrow{\sqrt{\mathrm{P}(Y)}}$

Proof.

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

$$= \sum_{j=1}^{|X|} \frac{\mathrm{P}(x_j,y_i)}{\sqrt{\mathrm{P}(y_i)}} = \sqrt{\mathrm{P}(y_i)}$$

□

Claim 3.6. $\|\overrightarrow{\sqrt{\mathrm{P}(X)}} \cdot \mathrm{M}\|_2 > 1 - 8\alpha \log\left((\alpha'+\delta)/\alpha\right)$

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

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

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

$$\begin{aligned}\|\overrightarrow{\sqrt{\mathrm{P}(X)}} \cdot \mathrm{M}\|_2 &= \|\overrightarrow{\sqrt{\mathrm{P}(X)}} \cdot \mathrm{M}\|_2 \cdot \|\overrightarrow{\sqrt{\mathrm{P}(Y)}^T}\|_2 \geq \overrightarrow{\sqrt{\mathrm{P}(X)}} \cdot \mathrm{M} \cdot \overrightarrow{\sqrt{\mathrm{P}(Y)}^T} \\ &= \overrightarrow{\sqrt{\mathrm{P}(X)}} \cdot \mathrm{M}' \cdot \overrightarrow{\sqrt{\mathrm{P}(Y)}^T} - \overrightarrow{\sqrt{\mathrm{P}(X)}} \cdot \widetilde{\mathrm{M}} \overrightarrow{\sqrt{\mathrm{P}(Y)}^T}\end{aligned}$$

Using Claim (3.5) we obtain

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

Therefore, using Equation (10) we obtain,

$$\begin{aligned}\|\overrightarrow{\sqrt{\mathrm{P}(X)}} \cdot \mathrm{M}\|_2 &\geq 1 - \overrightarrow{\sqrt{\mathrm{P}(X)}} \cdot \widetilde{\mathrm{M}} \overrightarrow{\sqrt{\mathrm{P}(Y)}^T} \\ &= 1 - \alpha \left(1 - 1/(c_1 \cdot \sqrt{d_X \cdot d_Y})\right) - (1-\alpha)(1 - 1/(c_0 \cdot \sqrt{d_X \cdot d_Y})) \\ &\geq 1 - \alpha - (1-\alpha)(1 - 1/(c_0 \cdot \sqrt{d_X \cdot d_Y})) \\ &= (1-\alpha)/(c_0 \cdot \sqrt{d_X \cdot d_Y})\end{aligned}$$

By the bound we have on c_0 in Lemma (3.3)

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

□

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 $\overrightarrow{\sqrt{\mathrm{P}(X)}} = \sum_{i=1}^{|X|} a_i u_i$, that is, $\overrightarrow{\sqrt{\mathrm{P}(X)}}$ 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:

$$\begin{aligned}\|\overrightarrow{\sqrt{\mathrm{P}(X)}} \cdot \mathrm{M}\|_2^2 &\leq a_1^2 + (1-\lambda)^2 \sum_{i=2}^{|X|} a_i^2 \\ &= (1 - (1-\lambda)^2) a_1^2 + (1-\lambda)^2 \sum_{i=1}^{|X|} a_i^2 \\ &= (1 - (1-\lambda)^2) a_1^2 + (1-\lambda)^2 = (2\lambda - \lambda^2) a_1^2 + (1-\lambda)^2\end{aligned}$$

Thus,

$$a_1^2 (2\lambda - \lambda^2) \geq \|\overrightarrow{\sqrt{\mathrm{P}(X)}} \cdot \mathrm{M}\|_2^2 - (1-\lambda)^2$$

We now obtain,

$$\begin{aligned}a_1 \geq a_1^2 &\geq \|\overrightarrow{\sqrt{\mathrm{P}(X)}} \cdot \mathrm{M}\|_2^2/(2\lambda - \lambda^2) - 1/(2\lambda - \lambda^2) + 1 \\ &\geq \left(1 - 8\alpha \log\left((\alpha' + \delta)/\alpha\right)\right)^2/(2\lambda - \lambda^2) - 1/(2\lambda - \lambda^2) + 1 \\ &\geq 1 - 16\alpha \log\left((\alpha' + \delta)/\alpha\right)/(2\lambda - \lambda^2)\end{aligned}$$

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

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

In the same way we can derive

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

□

Claim 3.7. $\delta \leq 16\alpha'$

Proof. Without loss of generality assume that

$$\frac{1}{d_X} \cdot \sum_{(x,y) \in S} \mathrm{P}(x) - \mathrm{Q}(x) \geq \frac{1}{d_Y} \cdot \sum_{(x,y) \in S} \mathrm{P}(y) - \mathrm{Q}(y).$$

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

Recall that

$$\sum_{x \in X} \frac{1}{\sqrt{|X|}} \sqrt{\mathrm{P}(x)} \geq 1 - 16\alpha \log\left((\alpha' + \delta)/\alpha\right)/(2\lambda - \lambda^2). \tag{11}$$

We will assume by way of contradiction that $\delta > 16\alpha'$ and obtain a contradiction to Equation (11). 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 $\sum_{x \in S'} \mathrm{P}(x) - \mathrm{Q}(x) \geq \delta$, we can assume in order to derive a contradiction to Equation (11) that $\sum_{x \in S'} \mathrm{P}(x) - \mathrm{Q}(x) = \delta$. Thus we can assume that P_X is as follows:

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

Thus

$$\begin{aligned} &\sum_{x \in X} \frac{1}{\sqrt{|X|}} \sqrt{\mathrm{P}(x)} = \alpha' \cdot \sqrt{1 + \delta/\alpha'} + (1 - \alpha')\sqrt{1 - \delta/(1 - \alpha')} \\ &\leq \alpha' \cdot \left(1 + \sqrt{\delta/\alpha'}\right) + (1 - \alpha')\left(1 - \delta/(2(1 - \alpha'))\right) \\ &= \sqrt{\delta \cdot \alpha'} + 1 - \delta/2 \end{aligned}$$

Since $\delta > 16\alpha'$, $\sqrt{\delta \cdot \alpha'} + 1 - \delta/2 \leq 1 - \delta/4$. Hence by Equation (11)

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

Since $0 \leq \lambda \leq 1$

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

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

$$\delta/\alpha < 64 \log\left((17/16)\delta/\alpha\right)/\lambda$$

That is

$$\delta/\alpha - 64 \log\left((17/16)\delta/\alpha\right)/\lambda < 0 \tag{12}$$

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

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

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

$$f(z) = z - 64 \log\left((17/16)z\right)/\lambda$$

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 (12).

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

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

Corollary 3.8.

$$\sum_{x \in X} \frac{1}{\sqrt{|X|}} \sqrt{\mathrm{P}(x)} \geq 1 - 16(\alpha/\lambda) \log\left(17\alpha'/\alpha\right)$$

$$\sum_{y \in Y} \frac{1}{\sqrt{|Y|}} \sqrt{\mathrm{P}(y)} \geq 1 - 16(\alpha/\lambda) \log\left(17\alpha'/\alpha\right)$$

Proof. By Lemma (3.4)

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

The corollary follows by the bound on δ in Claim (3.7), ($\delta \leq 16\alpha'$) and since $1 \leq 2 - \delta \leq 2$. □

We denote

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

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 $\mathrm{Q}(S_1), \mathrm{Q}(S_3)$ are small since $\mathrm{Q}(X_1)$ and $\mathrm{Q}(X_3)$ are small and that $\mathrm{Q}(S_2) \leq 40\alpha$

Claim 3.9. $\mathrm{Q}(S_1) \leq 400(\alpha/\lambda)\log(17\alpha'/\alpha)$

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

$$\mathrm{Q}(X_1) > 400(\alpha/\lambda)\log\left(17\alpha'/\alpha\right)$$

and obtain a contradiction to Corollary (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 $x \in X_1$ $\mathrm{P}(x)/\mathrm{Q}(x) < 1/2$, we can assume in order to derive a contradiction that for every $x \in X_1$ $\mathrm{P}(x)/\mathrm{Q}(x) = 1/2$. Denote

$$q := \mathrm{Q}(X_1).$$

Thus we can assume that P_X is as follows:

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

Thus

$$\begin{aligned} \sum_{x \in X} \frac{1}{\sqrt{|X|}}\sqrt{\mathrm{P}(x)} &= q \cdot 1/\sqrt{2} + (1-q)\sqrt{1 + \frac{q}{2(1-q)}} \\ &\leq q/\sqrt{2} + (1-q)(1 + \frac{q}{4(1-q)}) \\ &= q/\sqrt{2} + (1-q) + \frac{q}{4} \\ &< 1 - 0.04q \end{aligned}$$

Hence by Corollary (3.8)

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

Thus,

$$q < 400(\alpha/\lambda)\log\left(17\alpha'/\alpha\right).$$

□

Claim 3.10. $\mathrm{Q}(S_3) \leq 200(\alpha/\lambda)\log(17\alpha'/\alpha)$

Proof. Note that

$$\mathrm{Q}(S_3) \leq \mathrm{Q}(X_3).$$

We will assume by way of contradiction that $\mathrm{Q}(X_3) > 200(\alpha/\lambda)\log(17\alpha'/\alpha)$ and obtain a contradiction to Corollary (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 $x \in X_3$ $\mathrm{P}(x)/\mathrm{Q}(x) > 2$, we can assume in order to derive a contradiction that for every $x \in X_3$ $\mathrm{P}(x)/\mathrm{Q}(x) = 2$. Denote

$$q := \mathrm{Q}(X_3).$$

Thus we can assume that P_X is as follows:

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

Thus

$$\begin{aligned} &\sum_{x\in X} \frac{1}{\sqrt{|X|}}\sqrt{\mathrm{P}(x)} = q\cdot\sqrt{2} + (1-q)\sqrt{1-\frac{q}{1-q}} \\ &\leq q\sqrt{2} + (1-q)(1-\frac{q}{2(1-q)}) \\ &= q\sqrt{2} + (1-q) - \frac{q}{2} \\ &< 1 - 0.08q \end{aligned}$$

Hence by Corollary (3.8)

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

Thus,

$$q < 200(\alpha/\lambda)\log\left(17\alpha'/\alpha\right).$$

□

Claim 3.11. $\mathrm{Q}(S_2) \leq 40\alpha$

Proof. Denote

$$\widetilde{\mathrm{P}}(x,y) = \begin{cases} \mathrm{Q}(x)\mathrm{P}(y|x) & x \in X_2\ ; \\ \mathrm{Q}(x,y) & \text{otherwise.} \end{cases}$$

The proof will follow by combining Claim (3.12) and Claim (3.13). We will show that $\mathrm{Q}(S_2) \leq 4\widetilde{\mathrm{P}}(S_2) + 16\alpha$ and $\widetilde{\mathrm{P}}(S_2) \leq 2\mathrm{P}(S_2) \leq 2\alpha$.

Claim 3.12. $\widetilde{\mathrm{P}}(S_2) \leq 2\mathrm{P}(S_2)$

Proof. Since for every $x \in X_2$ $\mathrm{Q}(x) \leq 2\mathrm{P}(x)$

$$\widetilde{\mathrm{P}}(S_2) = \sum_{(x,y)\in S_2} \mathrm{Q}(x)\mathrm{P}(y|x) \leq 2 \sum_{(x,y)\in S_2} \mathrm{P}(x)\mathrm{P}(y|x) = 2\mathrm{P}(S_2)$$

□

Claim 3.13. $\mathrm{Q}(S_2) \leq 4\widetilde{\mathrm{P}}(S_2) + 16\alpha$

Proof. We will show that $\mathrm{D}(\widetilde{\mathrm{P}}\|\mathrm{Q}) \leq 2\epsilon$ and then use Lemma (2.6) to conclude the claim. By definition,

$$\begin{aligned}
\mathrm{D}(\widetilde{\mathrm{P}}\|\mathrm{Q}) &= \sum_{(x,y)\in X\times Y} \widetilde{\mathrm{P}}(x,y) \log\left(\frac{\widetilde{\mathrm{P}}(x,y)}{\mathrm{Q}(x,y)}\right) \\
&= \sum_{(x,y),\ x\in X_2} \mathrm{Q}(x)\mathrm{P}(y|x) \log\left(\frac{\mathrm{Q}(x)\mathrm{P}(y|x)}{\mathrm{Q}(x)\mathrm{Q}(y|x)}\right) + \sum_{(x,y),\ x\notin X_2} \widetilde{\mathrm{P}}(x,y) \log\left(\mathrm{Q}(x,y)/\mathrm{Q}(x,y)\right) \\
&= \sum_{(x,y),\ x\in X_2} \frac{\mathrm{Q}(x)}{\mathrm{P}(x)} \cdot \mathrm{P}(x)\mathrm{P}(y|x) \log\left(\frac{\mathrm{Q}(x)\mathrm{P}(y|x)}{\mathrm{Q}(x)\mathrm{Q}(y|x)}\right)
\end{aligned}$$

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

$$\mathrm{D}(\widetilde{\mathrm{P}}\|\mathrm{Q}) \leq 2 \sum_{(x,y),\ x\in X_2} \mathrm{P}(x,y) \log\left(\frac{\mathrm{P}(y|x)}{\mathrm{Q}(y|x)}\right) \leq 2\mathbb{E}_{X\sim \mathrm{P}_X} \mathrm{D}(\mathrm{P}_{Y|X}\|\mathrm{Q}_{Y|X}) \leq 4\epsilon$$

where the last inequality follows from the assumption on P. Applying Corollary (2.7) we obtain that $\mathrm{Q}(S_2) \leq 4\widetilde{\mathrm{P}}(S_2) + 16\alpha$ □

Combining Claim (3.12) and Claim (3.13) we obtain $\mathrm{Q}(S_2) \leq 8\mathrm{P}(S_2) + 16\alpha \leq 24\alpha$ □

We now prove Lemma (3.1). Recall that,

$$\mathrm{Q}(S) = \mathrm{Q}(S_1) + \mathrm{Q}(S_2) + \mathrm{Q}(S_3)$$

Combining Claim (3.9), Claim (3.11) and Claim (3.10) we obtain:

$$\begin{aligned}
\mathrm{Q}(S) &\leq 400(\alpha/\lambda)\log\left(17\alpha'/\alpha\right) + 40\alpha + 200(\alpha/\lambda)\log\left(17\alpha'/\alpha\right) \\
&\leq 600(\alpha/\lambda)\log\left(17\alpha'/\alpha\right) + 40\alpha \\
&\leq 640(\alpha/\lambda)\log\left(17\alpha'/\alpha\right)
\end{aligned}$$

That is

$$\alpha'/\alpha \leq (640/\lambda)\log\left(17\alpha'/\alpha\right)$$

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

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

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 $z > 640/(\lambda \ln 2)$ □

4 Proof of Main Results

Recall the notations from Section (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 $j \neq i$.

4.1 General Games

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

Lemma 4.1. *For general games and the event W,*

$$\begin{aligned}\mathbb{E}_{i\in[n-k]}\big(\ & \mathbb{E}_{X^k,Y^k,A^k,M^{-i}|W}\mathbb{E}_{X_i|X^k,Y^k,A^k,M^{-i},W}\mathrm{D}(\mathrm{P}_{Y_i|X_i,X^k,Y^k,A^k,M^{-i},W}\|\mathrm{P}_{Y_i|X_i}) + \\ & \mathbb{E}_{X^k,Y^k,A^k,M^{-i}|W}\mathbb{E}_{Y_i|X^k,Y^k,A^k,M^{-i},W}\mathrm{D}(\mathrm{P}_{X_i|Y_i,X^k,Y^k,A^k,M^{-i},W}\|\mathrm{P}_{X_i|Y_i})\ \big) \\ & \leq \frac{1}{n-k}(-\log(\Pr[W]) + k\log s)\end{aligned}$$

where s is the size of the answers set.

Proof. By [24] Claim (5.3) in the proof of Lemma (4.2) we obtain that for every $x^k \in X^k, y^k \in Y^k, a^k \in A^k$,

$$\begin{aligned}\mathbb{E}_{i\in[n-k]}\big(\ &\mathbb{E}_{M^{-i}|x^k,y^k,a^k,W}\mathbb{E}_{X_i|x^k,y^k,a^k,M^{-i},W}\mathrm{D}(\mathrm{P}_{Y_i|X_i,x^k,y^k,a^k,M^{-i},W}\|\mathrm{P}_{Y_i|X_i})+\\ &\mathbb{E}_{M^{-i}|x^k,y^k,a^k,W}\mathbb{E}_{Y_i|x^k,y^k,a^k,M^{-i},W}\mathrm{D}(\mathrm{P}_{X_i|Y_i,x^k,y^k,a^k,M^{-i},W}\|\mathrm{P}_{X_i|Y_i})\ \big)\\ &\le \frac{1}{n-k}(-\log(\Pr[W|X^k=x^k,Y^k=y^k])+\log(1/\Pr[A^k=a^k|X^k=x^k,Y^k=y^k,W]))\end{aligned}$$

Using this, we obtain:

$$\begin{aligned}&\mathbb{E}_{X^k,Y^k,A^k|W}\mathbb{E}_{i\in[n-k]}\big(\ \mathbb{E}_{M^{-i}|X^k,Y^k,A^k,W}\mathbb{E}_{X_i|X^k,Y^k,A^k,M^{-i},W}\mathrm{D}(\mathrm{P}_{Y_i|X_i,X^k,Y^k,A^k,M^{-i},W}\|\mathrm{P}_{Y_i|X_i})+\\ &\mathbb{E}_{M^{-i}|X^k,Y^k,A^k,W}\mathbb{E}_{Y_i|X^k,Y^k,A^k,M^{-i},W}\mathrm{D}(\mathrm{P}_{X_i|Y_i,X^k,Y^k,A^k,M^{-i},W}\|\mathrm{P}_{X_i|Y_i})\ \big)\\ &\le \mathbb{E}_{X^k,Y^k,A^k|W}\cdot\frac{1}{n-k}\Big(-\log(\Pr[W|X^k,Y^k])+\log(1/\Pr[A^k|X^k,Y^k,W])\Big)\\ &\le \frac{1}{n-k}\cdot\log\left(\mathbb{E}_{X^k,Y^k|W}\cdot\frac{1}{\Pr[W|X^k,Y^k]}\right)+\frac{1}{n-k}\cdot\mathbb{E}_{X^k,Y^k,A^k|W}\log(1/\Pr[A^k|X^k,Y^k,W])\\ &\le \frac{-\log(\Pr[W])}{n-k}+\frac{1}{n-k}\cdot\mathbb{E}_{X^k,Y^k|W}\mathbb{E}_{A^k|X^k,Y^k,W}\log(1/\Pr[A^k|X^k,Y^k,W]) \qquad (13)\\ &= \frac{-\log(\Pr[W])}{n-k}+\frac{1}{n-k}\mathbb{E}_{X^k,Y^k|W}\mathrm{H}(\mathrm{P}_{A^k|X^k,Y^k,W}) \qquad (14)\end{aligned}$$

We use the trivial bound on the size of the support, namely, for every x^k, y^k we can bound $|\mathrm{supp}(\mathrm{P}_{A^k|X^k=x^k,Y^k=y^k,W})| \le |\mathrm{supp}(\mathrm{P}_{A^k})| \le s^k$ where s is the size of the answers set. Using Fact 2.3 we can conclude the lemma. □

Lemma 4.2 (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* $\Pr(W) \ge 2^{-\epsilon'(n-k)/4+k\log s}$ *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*

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

Proof. For every $i\in[n-k]$ and x^k, y^k, a^k, m^{-i} , we will use a strategy for the game $G(\mathrm{P}_{X^n,Y^n}, V^{\otimes n})$ to obtain a strategy for the game

$$G(\mathrm{P}_{X_iY_i|X^k=x^k,Y^k=y^k,A^k=a^k,M^{-i}=m^{-i},W}, V).$$

Fix any strategy, f_a, f_b, for the game $\mathrm{G}(\mathrm{P}_{X^nY^n}, V^{\otimes n})$, and apply the following to obtain a strategy for

$$G(\mathrm{P}_{X_iY_i|X^k=x^k,Y^k=y^k,A^k=a^k,M^{-i}=m^{-i},W}, V):$$

Algorithm 4.3. *Protocol for* $G(\mathrm{P}_{X_iY_i|X^k=x^k,Y^k=y^k,A^k=a^k,M^{-i}=m^{-i},W}, V)$ *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* $\mathrm{P}_{X_iY_i|X^k=x^k,Y^k=y^k,A^k=a^k,M^{-i}=m^{-i},W}$*. 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* $\mathrm{P}_{X^{n-k}|X^k=x^k,Y^k=y^k,A^k=a^k,M^{-i}=m^{-i},W,X_i=x}$*. 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* $\mathrm{P}_{Y^{n-k}|X^k=x^k,Y^k=y^k,A^k=a^k,M^{-i}=m^{-i},W,Y_i=y}$*. 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 4.4. *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* $\mathrm{P}_{X^m,Y^m|X^k=x^k,Y^k=y^k,A^k=a^k,X_i=x,Y_i=y,M^{-i}=m^{-i},W}$ *since conditioned on* X_i *(or* Y_i*) this distribution is a product distribution. Hence, the joint distribution of* x^n, y^n *is*

$$\mathrm{P}_{X^n,Y^n|X^k=x^k,Y^k=y^k,A^k=a^k,M^{-i}=m^{-i}}$$

Remark 4.5. *Notice that since Remark 4.4 holds, the probability of winning the game*

$$G(\mathrm{P}_{X_iY_i|X^k=x^k,Y^k=y^k,A^k=a^k,M^{-i}=m^{-i},W}, V)$$

is

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

Remark 4.6. *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*

$$G(\mathrm{P}_{X_iY_i|X^k=x^k,Y^k=y^k,A^k=a^k,M^{-i}=m^{-i},W}, V)$$

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 4.1 for a fixed strategy f_a, f_b for $\mathrm{G}(\mathrm{P}_{X^nY^n}, V^{\otimes n})$,

$$\begin{aligned}\mathbb{E}_{i\in[n-k]}(\ &\mathbb{E}_{X^k,Y^k,A^k,M^{-i}|W}\mathbb{E}_{X_i|X^k,Y^k,A^k,M^{-i},W}\mathrm{D}(\mathrm{P}_{Y_i|X_i,X^k,Y^k,A^k,M^{-i},W}\|\mathrm{P}_{Y_i|X_i})+\\ &\mathbb{E}_{X^k,Y^k,A^k,M^{-i}|W}\mathbb{E}_{Y_i|X^k,Y^k,A^k,M^{-i},W}\mathrm{D}(\mathrm{P}_{X_i|Y_i,X^k,Y^k,A^k,M^{-i},W}\|\mathrm{P}_{X_i|Y_i})\ \ )\\ &\leq \frac{1}{n-k}(-\log(\Pr[W])+k\log s)\end{aligned}$$

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

$$\begin{aligned}\mathbb{E}_{i\in[n-k]}(\ &\mathbb{E}_{X^k,Y^k,A^k,M^{-i}|W}\mathbb{E}_{X_i|X^k,Y^k,A^k,M^{-i},W}\mathrm{D}(\mathrm{P}_{Y_i|X_i,X^k,Y^k,A^k,M^{-i},W}\|\mathrm{P}_{Y_i|X_i})+\\ &\mathbb{E}_{X^k,Y^k,A^k,M^{-i}|W}\mathbb{E}_{Y_i|X^k,Y^k,A^k,M^{-i},W}\mathrm{D}(\mathrm{P}_{X_i|Y_i,X^k,Y^k,A^k,M^{-i},W}\|\mathrm{P}_{X_i|Y_i})\ \ ) \leq \epsilon'/4\end{aligned}$$

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 $i \in [n-k]$,

$$\mathbb{E}_{X^k,Y^k,A^k,M^{-i}|W} \Pr(\neg W_i | X^k, Y^k, A^k, M^{-i}, W) < \epsilon'/2.$$

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:

$$\begin{aligned}&\mathbb{E}_{X_i|X^k=x^k,Y^k=y^k,A^k=a^k,M^{-i}=m^{-i},W} \mathrm{D}(\mathrm{P}_{Y_i|X_i,X^k=x^k,Y^k=y^k,A^k=a^k,M^{-i}=m^{-i},W} \| \mathrm{P}_{Y_i|X_i}) + \\ &\mathbb{E}_{Y_i|X^k=x^k,Y^k=y^k,A^k=a^k,M^{-i}=m^{-i},W} \mathrm{D}(\mathrm{P}_{X_i|Y_i,X^k=x^k,Y^k=y^k,A^k=a^k,M^{-i}=m^{-i},W} \| \mathrm{P}_{X_i|Y_i}) \leq \epsilon'/2 \end{aligned} \quad (15)$$

$$\Pr(\neg W_i | X^k = x^k, Y^k = y^k, A^k = a^k, M^{-i} = m^{-i}, W) < \epsilon' \quad (16)$$

For the strategy f_a, f_b and for x^k, y^k, a^k, i, m^{-i} for which both Equation (15) and Equation (16) hold consider the protocol suggested in Algorithm 4.3. Recall that by Remark 4.6 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 (16)

$$\mathrm{P}_{X_i,Y_i|X^k=x^k,Y^k=y^k,A^k=a^k,M^{-i}=m^{-i},W}(R) < \epsilon'. \quad (17)$$

Combining Equation (17) with Equation (15), we can apply Lemma 3.1 to obtain $\mathrm{P}_{X_i,Y_i}(R) < \epsilon$. The provers can play h_a, h_b as a strategy for $\mathrm{G}(\mathrm{P}_{X_i,Y_i}, V)$ and err only on questions in R. Since $\mathrm{P}_{X_i,Y_i}(R) < \epsilon$, the value of $\mathrm{G}(\mathrm{P}_{X_i,Y_i}, V) > 1-\epsilon$ and since $\mathrm{P}_{X_i,Y_i} = \mathrm{P}_{XY}$, the value of $\mathrm{G}(\mathrm{P}_{XY}, V) > 1-\epsilon$ which is a contradiction. □

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

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

Proof. (Of Theorem 1): 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 4.2 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[3]

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

Thus, for $k \leq n\epsilon'/(16\log s)$ we can apply Lemma 4.2 to obtain that there is $i \notin T$ for which $\Pr(W_i|W) \leq 1 - \epsilon'/2$ therefore,

$$\Pr(W_i \wedge W) = \Pr(W) \cdot \Pr(W_i|W) \leq (1 - \epsilon'/2)^{k+1}$$

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:

$$\begin{aligned}\Pr(W_1 \wedge \ldots \wedge W_n) &\leq \Pr(\bigwedge_{i \in T} W_i) \leq (1 - \epsilon'/2)^{n\epsilon'/(16\log s)} \\ &\leq (1 - (\epsilon')^2/32)^{n/\log s}\end{aligned}$$

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 $x \leq 1$ □

4.2 Projection Games

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

Lemma 4.7 (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 (4.7)): 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

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

[3]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.

Claim 4.8. *Given the event W, with probability of at least 80% over x^k, y^k (that are chosen according to the conditional distribution $\mathrm{P}_{X^kY^k|W}$) both equations hold:*

$$\mathbb{E}_{i\notin T}\Pr(W_i|W, X^k = x^k, Y^k = y^k) > 1 - 10^{-7}\epsilon' \tag{18}$$

$$\Pr(W|X^k = x^k, Y^k = y^k) \geq 1 - 10^{-11}\lambda \tag{19}$$

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 $\mathrm{P}_{X^kY^k}$,

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

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

$$\begin{aligned}&\sum_{(x^k,y^k)\in E_1}\Pr(X^k = x^k, Y^k = y^k)\Pr(W|X^k = x^k, Y^k = y^k)\\ &< (1 - 10^{-11}\lambda)\cdot\sum_{(x^k,y^k)\in E_1}\Pr(X^k = x^k, Y^k = y^k) < 0.1\cdot\Pr(W).\end{aligned}$$

Hence,

$$\sum_{(x^k,y^k)\in E_1}\Pr(X^k = x^k, Y^k = y^k|W) < 0.1 \tag{20}$$

Notice that

$$\begin{aligned}\mathbb{E}_{i\notin T}\Pr(W_i|W) &= \mathbb{E}_{i\notin T}\Pr(W_i \wedge W)/\Pr(W)\\ &= \mathbb{E}_{i\notin T}\sum_{x^k\in X^k, y^k\in Y^k}\frac{\Pr(X^k = x^k, Y^k = y^k)\Pr(W|X^k = x^k, Y^k = y^k)}{\Pr(W)}\cdot\\ &\quad\Pr(W_i|X^k = x^k, Y^k = y^k, W)\\ &= \mathbb{E}_{i\notin T}\sum_{x^k\in X^k, y^k\in Y^k}\Pr(X^k = x^k, Y^k = y^k|W)\Pr(W_i|X^k = x^k, Y^k = y^k, W)\\ &= \sum_{x^k\in X^k, y^k\in Y^k}\Pr(X^k = x^k, Y^k = y^k|W)\mathbb{E}_{i\notin T}\Pr(W_i|X^k = x^k, Y^k = y^k, W)\end{aligned}$$

Denote

$$E_2 = \{(x^k, y^k)|\mathbb{E}_{i\notin T}\Pr(W_i|W, X^k = x^k, Y^k = y^k) < 1 - 10^{-7}\epsilon'\}.$$

Thus,

$$\begin{aligned}1 - 10^{-8}\epsilon' &< \mathbb{E}_{i\notin T}\Pr(W_i|W)\\ &< (1 - 10^{-7}\epsilon')\cdot\sum_{(x^k,y^k)\in E_2}\Pr(X^k = x^k, Y^k = y^k|W) + \sum_{(x^k,y^k)\notin E_2}\Pr(X^k = x^k, Y^k = y^k|W))\\ &= 1 - 10^{-7}\epsilon'\cdot\sum_{(x^k,y^k)\in E_2}\Pr(X^k = x^k, Y^k = y^k|W)\end{aligned}$$

Thus we obtain that

$$\sum_{(x^k,y^k)\in E_2} \Pr(X^k = x^k, Y^k = y^k|W) < 0.1. \tag{21}$$

By Equation (20) and Equation (21) 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 $\mathrm{P}_{X^kY^k|W}$, both Equation (18) and Equation (19) hold. □

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 4.9. *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

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

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

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

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

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

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

(where for $b^k \in B^k$, $f_{x^ky^k}(b^k)$ 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 (4.10), 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 $\mathrm{G}(\mathrm{P}_{XY}, V)$ 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

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

where $(x^{n-k}, y^{n-k}) \in E'$ if and only if for every $i \in \{1, \ldots, n-k\}$, $\mathrm{P}_{XY}(x_i, y_i) > 0$. 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 4.10. *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 $d = d_X d_Y$ regular graph with second normalized eigenvalue of $(1-\lambda)^2$.

Definition 4.11 (Edge Expansion)**.** *The edge expansion $h(G)$ of a graph $G = (V, E)$ is defined as*

$$h(G) = \min_{S \subseteq V, 1 \le |S| \le |V|/2} \frac{\partial(S)}{|S|}$$

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 4.12. *For a d regular expander graph G with second eigenvalue $1-\lambda$,*

$$h(G) \ge \frac{1}{2}(d - d(1-\lambda)) = \frac{1}{2}d\lambda.$$

Hence,

$$h(G^2) \ge \frac{1}{2}(d - d(1-\lambda)^2) \ge \frac{1}{2}d\lambda$$

See proof in [4], [5], [6]

By the expansion property of G^2 the number of edges between[4] X' and X'' is at least $|X'|h(G^2)$. By Fact (4.12) we obtain that the number of edges in G^2 between X' and X'' is at least

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

□

[4]We assume without loss of generality that $|X'| \le 1/2|X^{n-k}|$ otherwise we do the same argument on X''.

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 lease $(1/4)|X'|d_X\lambda$. Recall that $|X'|/|X^{n-k}| \geq 1/4$, therefore, the probability of wining 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 (4.8) and Claim (4.9), with probability of at least 80% over x^k, y^k that are taken according to the distribution $\mathrm{P}_{X^kY^k|W}$ there exists a^k such that the following equations hold:

$$\mathbb{E}_{i\notin T}\Pr(W_i|W, X^k = x^k, Y^k = y^k) > 1 - 10^{-7}\epsilon' \tag{22}$$
$$\Pr(W|X^k = x^k, Y^k = y^k) \geq 1 - 10^{-11}\lambda \tag{23}$$
$$\Pr(A^k = a^k|X^k = x^k, Y^k = y^k) \geq 1/2 \tag{24}$$

Claim 4.13. *For every x^k, y^k, a^k that satisfy Equation (23) and Equation (24)*

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

Proof. Denote the event that the provers loose on the last k coordinates by $\neg W$. Since,

$$\begin{aligned}\Pr(\neg W|X^k = x^k, Y^k = y^k) &= \sum_{\alpha^k\in A^k}\Pr(A^k = \alpha^k|X^k = x^k, Y^k = y^k)\Pr(\neg W|X^k = x^k, Y^k = y^k, A^k = \alpha^k)\\ &\geq \Pr(A^k = a^k|X^k = x^k, Y^k = y^k)\Pr(\neg W|X^k = x^k, Y^k = y^k, A^k = a^k)\\ &\geq (1/2)\Pr(\neg W|X^k = x^k, Y^k = y^k, A^k = a^k)\end{aligned}$$

By combining Equation (23) with Equation (24) we obtain

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

□

Claim 4.14. *With probability of at least 70% over x^k, y^k that are taken according to the distribution $\mathrm{P}_{X^kY^k|W}$, 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

$$\begin{aligned}&\Pr(\neg W_i|W, X^k = x^k, Y^k = y^k) = \sum_{a^k\in A^k}\Pr(\neg W_i \wedge (A^k = a^k)|W, X^k = x^k, Y^k = y^k)\\ &= \sum_{a^k\in A^k}\Pr(A^k = a^k|W, X^k = x^k, Y^k = y^k)\cdot\Pr(\neg W_i|W, X^k = x^k, Y^k = y^k, A^k = a^k)\\ &= \sum_{a^k\in A^k}\frac{\Pr(A^k = a^k|X^k = x^k, Y^k = y^k)\cdot\Pr(W|X^k = x^k, Y^k = y^k, A^k = a^k)}{\Pr(W|X^k = x^k, Y^k = y^k)}\cdot\\ &\quad\Pr(\neg W_i|W, X^k = x^k, Y^k = y^k, A^k = a^k)\end{aligned} \tag{25}$$

With probability of at least 80% over x^k, y^k that are taken according to the distribution $\mathrm{P}_{X^kY^k|W}$ there exists a^k for which both Equation (24) and Claim (4.13) hold. Thus, using Equation (25), Claim (4.13) and Equation (24), for those x^k, y^k, a^k and for every $i \notin T$:

$$\begin{aligned}&\Pr(\neg W_i|W, X^k = x^k, Y^k = y^k) \geq (1/2)\cdot(1 - 10^{-10}\lambda)\cdot\Pr(\neg W_i|W, X^k = x^k, Y^k = y^k, A^k = a^k)\\ &\geq (1/4)\cdot\Pr(\neg W_i|W, X^k = x^k, Y^k = y^k, A^k = a^k)\end{aligned}$$

Therefore,

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

Hence the claim follows. □

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

$$\mathbb{E}_{i\notin T}\Pr(W_i|W,X^k=x^k,Y^k=y^k,A^k=a^k) > 1-10^{-6}\epsilon' \tag{26}$$
$$\Pr(W|X^k=x^k,Y^k=y^k)\geq 1-10^{-11}\lambda \tag{27}$$
$$\Pr(W|X^k=x^k,Y^k=y^k,A^k=a^k)\geq 1-10^{-10}\lambda$$
$$\Pr(A^k=a^k|X^k=x^k,Y^k=y^k)\geq 1/2$$

By [24] Claim (5.3) in the proof of Lemma (4.2) we obtain that for every $x^k\in X^k, y^k\in Y^k, a^k\in A^k$,

$$\begin{aligned}\mathbb{E}_{i\in[n-k]}(\ \ &\mathbb{E}_{M^{-i}|x^k,y^k,a^k,W}\mathbb{E}_{X_i|x^k,y^k,a^k,M^{-i},W}\mathrm{D}(\mathrm{P}_{Y_i|X_i,x^k,y^k,a^k,M^{-i},W}\|\mathrm{P}_{Y_i|X_i})+ \\ &\mathbb{E}_{M^{-i}|x^k,y^k,a^k,W}\mathbb{E}_{Y_i|x^k,y^k,a^k,M^{-i},W}\mathrm{D}(\mathrm{P}_{X_i|Y_i,x^k,y^k,a^k,M^{-i},W}\|\mathrm{P}_{X_i|Y_i})\ \ ) \\ &\leq \frac{1}{n-k}(-\log(\Pr[W|X^k=x^k,Y^k=y^k])+\log(1/\Pr[A^k=a^k|X^k=x^k,Y^k=y^k,W]))\end{aligned} \tag{28}$$

Since,

$$\begin{aligned}\Pr[A^k=a^k|X^k=x^k,Y^k=y^k,W] &= \frac{\Pr[A^k=a^k|X^k=x^k,Y^k=y^k]\cdot\Pr[W|X^k=x^k,Y^k=y^k,A^k=a^k]}{\Pr[W|X^k=x^k,Y^k=y^k]} \\ &\geq \Pr[A^k=a^k|X^k=x^k,Y^k=y^k]\cdot\Pr[W|X^k=x^k,Y^k=y^k,A^k=a^k] \\ &\geq 1/2\cdot(1-10^{-10}\lambda)\end{aligned}$$

and by applying both Equation (27) and the assumption

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

we obtain:

$$\text{Equation (28)} \leq (10^{-3}\cdot\epsilon'\cdot\lambda)\left(-\log(1-10^{-11}\lambda)-\log(0.5\cdot(1-10^{-10}\lambda))\right) < 10^{-2}\epsilon'$$

where the last inequality follows by using that $-\log(1-x)\leq 1$ for $x<1/2$ (recall that $\lambda<1$). Since

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

By Equation (26) we can fix $i\in[n-k]$ and $m^{-i}\in M^{-i}$ for which

$$\Pr(W_i|W,X^k=x^k,Y^k=y^k,A^k=a^k,M^{-i}=m^{-i})>1-\epsilon' \tag{29}$$

and

$$\begin{aligned}&\mathbb{E}_{X_i|X^k=x^k,Y^k=y^k,A^k=a^k,M^{-i}=m^{-i},W}\mathrm{D}(\mathrm{P}_{Y_i|X_i,X^k=x^k,Y^k=y^k,A^k=a^k,M^{-i}=m^{-i},W}\|\mathrm{P}_{Y_i|X_i})+ \\ &\mathbb{E}_{Y_i|X^k=x^k,Y^k=y^k,A^k=a^k,M^{-i}=m^{-i},W}\mathrm{D}(\mathrm{P}_{X_i|Y_i,X^k=x^k,Y^k=y^k,A^k=a^k,M^{-i}=m^{-i},W}\|\mathrm{P}_{X_i|Y_i})<10^{-1}\epsilon'\end{aligned} \tag{30}$$

For the strategy f_a, f_b and for x^k, y^k, a^k, i, m^{-i} for which both Equation (29) and Equation (30) hold, consider the protocol suggested in Algorithm 4.3. Recall that by Remark 4.6 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 (29)

$$\mathrm{P}_{X_i,Y_i|X^k=x^k,Y^k=y^k,A^k=a^k,M^{-i}=m^{-i},W}(R) < \epsilon'. \tag{31}$$

Combining Equation (31) with Equation (30), we can apply Lemma 3.1 to obtain $\mathrm{P}_{X_i,Y_i}(R) < \epsilon$. The provers can play h_a, h_b as a strategy for $\mathrm{G}(\mathrm{P}_{X_i,Y_i}, V)$ and err only on questions in R. Since $\mathrm{P}_{X_i,Y_i}(R) < \epsilon$, the value of $\mathrm{G}(\mathrm{P}_{X_i,Y_i}, V) > 1-\epsilon$ and since $\mathrm{P}_{X_i,Y_i} = \mathrm{P}_{XY}$, the value of $\mathrm{G}(\mathrm{P}_{XY}, V) > 1-\epsilon$ which is a contradiction. □

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

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

Proof Sketch By inductive application of Lemma (4.7) as in the proof of Theorem (1) we can reduce the value of the game from $1-\epsilon$ to $1-\Omega(\lambda)$ by $O(1/(\epsilon \cdot \mathrm{poly}(\lambda)))$ repetitions. Then by applying Rao's bound [23] we can further reduce the value of the game to any constant by $O(1/\mathrm{poly}(\lambda))$ repetitions of this protocol (that is, a total number of $O(1/(\epsilon \cdot \mathrm{poly}(\lambda)))$ repetitions.

Note that in order to apply Lemma (4.7) we need n to be large enough. Nevertheless, as suggested in [23], 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 (2)): By the observation above, we may assume throughout the proof that[5]

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

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

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

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

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

If

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

[5] where $\epsilon' = 10^{-6}\epsilon\lambda/\log(2/\lambda)$

then we are done. Otherwise

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

where we used the inequality $(1-x) \geq 2^{-2x}$ for $0 \leq x \leq 1/2$. In order to use Lemma 4.7 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

$$2^{-2(k+1)\epsilon \cdot 10^{-14} \cdot \lambda / \log \frac{2}{\lambda}} \geq 2^{-10^{-12}\lambda}.$$

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

$$\left(1 - 10^{-14} \cdot \epsilon \cdot \lambda / \log \frac{2}{\lambda}\right)^{\log(2/\lambda)\epsilon^{-1}} \leq 1 - 10^{-15} \cdot \lambda \tag{32}$$

We now state Rao's theorem ([23] 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 (32) with Rao's theorem, we obtain that the value of $G^{\otimes n}$ is at most $(1 - \Omega(\lambda^2))^{n \cdot 10^{-4}(\epsilon')\lambda} \leq (1 - \epsilon \cdot \mathrm{poly}(\lambda))^n$ □

References

[1] Sanjeev Arora, Boaz Barak, David Steurer. A $2^{n^{poly(1/\epsilon)}}$-Time Algorithm for Unique Games. Manuscript, 2010

[2] Sanjeev Arora, Russell Impagliazzo, William Matthews, David Steurer. Improved Algorithms for Unique Games via Divide and Conquer. Manuscript 2010

[3] Sanjeev Arora, Subhash Khot, Alexandra Kolla, David Steurer, Madhur Tulsiani, Nisheeth K. Vishnoi. Unique Games on Expanding Constraint Graphs are Easy: extended abstract. STOC 2008: 21-28

[4] Noga Alon and Vitali D. Milman. λ_1, isoperimetric inequalities for graphs, and superconcentrators. Journal of Combinatorial Theory. Series B. 38: 73–88 (1985)

[5] J. Dodziuk. Difference equations, isoperimetric inequality and transience of certain random walks. Trans. Amer. Math. Soc., 284(2):787794 (1984).

[6] Noga Alon. Eigenvalues and expanders. Combinatorica, 6(2):8396 (1986). Theory of Computing (Singer Island, FL, 1984).

[7] Boaz Barak, Mark Braverman, Xi Chen, Anup Rao. How to Compress Interactive Communication. STOC 2010

[8] Mihir Bellare, Oded Goldreich, Madhu Sudan. Free Bits, PCPs, and Nonapproximability-Towards Tight Results. *SIAM J. Comput.* 27(3): 804-915 (1998) (preliminary version in FOCS 1995)

[9] Boaz Barak, Moritz Hardt, Ishay Haviv, Anup Rao, Oded Regev, David Steurer. Rounding Parallel Repetitions of Unique Games. FOCS 2008: 374-383

[10] Boaz Barak, Anup Rao, Ran Raz, Ricky Rosen, Ronen Shaltiel. Strong Parallel Repetition Theorem for Free Projection Games. APPROX-RANDOM 2009: 352-365

[11] Richard Cleve, Peter Høyer, Benjamin Toner, John Watrous. Consequences and Limits of Nonlocal Strategies. CCC 2004: 236-249

[12] Uriel Feige. A Threshold of ln n for Approximating Set Cover. *J. ACM* 45(4): 634-652 (1998) (preliminary version in STOC 1996)

[13] Uriel Feige, Guy Kindler, Ryan O'Donnell. Understanding Parallel Repetition Requires Understanding Foams. CCC 2007: 179-192

[14] Uriel Feige, László Lovász. Two-Prover One-Round Proof Systems: Their Power and Their Problems STOC 1992: 733-744

[15] Uriel Feige, Oleg Verbitsky: Error Reduction by Parallel Repetition - A Negative Result. *Combinatorica* 22(4): 461-478 (2002) (preliminary version in CCC 1996)

[16] Johan Håstad. Some Optimal Inapproximability Results. *J. ACM* 48(4): 798-859 (2001) (preliminary version in STOC 1997)

[17] Thomas Holenstein. Parallel Repetition: Simplifications and the No-Signaling Case. *Theory of Computing* 5(1): 141-172 (2009) (preliminary version in STOC 2007)

[18] Subhash Khot. On The Power Of Unique 2-Prover 1-Round Games. STOC 2002: 767-775

[19] Alexandra Kolla. Spectral Algorithms for Unique Games. CCC 2010

[20] Guy Kindler, Ryan O'Donnell, Anup Rao and Avi Wigderson. Rounding Schemes and Cubical Tilings with Sphere-Like Surface Area. FOCS 2008: 189-198

[21] Konstantin Makarychev, Yury Makarychev. How to Play Unique Games on Expanders. Manuscript 2009

[22] Itzhak Parnafes, Ran Raz, Avi Wigderson. Direct Product Results and the GCD Problem, in Old and New Communication Models. STOC 1997: 363-372

[23] Anup Rao. Parallel Repetition in Projection Games and a Concentration Bound. STOC 2008: 1-10

[24] Ran Raz. A Parallel Repetition Theorem. *SIAM J. Comput.* 27(3): 763-803 (1998) (preliminary version in STOC 1995)

[25] Ran Raz. A Counterexample to Strong Parallel Repetition. FOCS 2008: 369-373

[26] Ran Raz. Parallel Repetition of two player games. CCC 2010

[27] Muli Safra, Oded Schwartz. On Parallel-Repetition, Unique-Game and Max-Cut. Manuscript 2007

A Analysis of the Boundaries in the Minimization Argument in the Proof of Lemma (3.2)

In this section, we consider the case that for some (x_0, y_0), $\mathrm{P}(x_0, y_0) = 0$ or $\mathrm{P}(x_0, y_0) = 1$, in the proof of Lemma (3.2) (see Footnote 1).

Case 1: There exists (x_0, y_0) for which $\mathrm{P}(x_0, y_0) = 1$; then it must be that $\mathrm{P}(x_0) = 1$ and $\mathrm{P}(y_0) = 1$, thus, Lemma (3.2) holds by taking $c_0 = 0$ and $c_1 = 1$ or vice-versa.

Case 2: There exists a distribution P that minimizes

$$\mathbb{E}_{Y\sim \mathrm{P}_Y} \mathrm{D}(\mathrm{P}_{X|Y} \| \mathrm{Q}_{X|Y}) + \mathbb{E}_{X\sim \mathrm{P}_X} \mathrm{D}(\mathrm{P}_{Y|X} \| \mathrm{Q}_{Y|X}) \tag{33}$$

(under the constraints), for which there exists (x_0, y_0) such that $\mathrm{P}(x_0, y_0) = 0$ and $\mathrm{P}(x_0) > 0$. Without loss of generality, assume $(x_0, y_0) \in S$. Let $(x_1, y_1) \in S$ be such that $\mathrm{P}(x_1, y_1) > 0$. Denote $\tau = \mathrm{P}(x_1, y_1)$.

By definition,

$$\mathbb{E}_{X\sim \mathrm{P}_X} \mathrm{D}(\mathrm{P}_{Y|X} \| \mathrm{Q}_{Y|X}) = \sum_{(x,y)\in X\times Y} \mathrm{P}(x,y) \log \left(\frac{\mathrm{P}(x,y)/\mathrm{P}(x)}{\mathrm{Q}(x,y)/\mathrm{Q}(x)} \right)$$

and similarly,

$$\mathbb{E}_{Y\sim \mathrm{P}_Y} \mathrm{D}(\mathrm{P}_{X|Y} \| \mathrm{Q}_{X|Y}) = \sum_{(x,y)\in X\times Y} \mathrm{P}(x,y) \log \left(\frac{\mathrm{P}(x,y)/\mathrm{P}(y)}{\mathrm{Q}(x,y)/\mathrm{Q}(y)} \right)$$

Note that the derivative of Equation (33) in the variable $\mathrm{P}(x_0, y_0)$ at the point $\mathrm{P}(x_0, y_0) = 0$ is $-\infty$, while the derivative of Equation (33) in the variable $\mathrm{P}(x_1, y_1)$ at the point $\mathrm{P}(x_1, y_1) = \tau$ is

finite. Thus, if we move a tiny mass from $\mathrm{P}(x_1, y_1)$ to $\mathrm{P}(x_0, y_0)$, the new distribution would decrease Equation (33). This is a contradiction to the assumption that P minimizes Equation (33).

A symmetric argument shows that the case where there exists (x_0, y_0) such that $\mathrm{P}(x_0, y_0) = 0$ and $\mathrm{P}(y_0) > 0$ does not minimize Equation (33).

Case 3: There exists a distribution P that minimizes Equation (33) (under the constraints), such that for all (x_0, y_0) for which $\mathrm{P}(x_0, y_0) = 0$, we have

$$\mathrm{P}(y_0) = 0, \quad \mathrm{P}(x_0) = 0.$$

Consider the set of all the points that satisfy the conditions above and denote this set by T. That is,

$$T := \{(x_0, y_0) \mid \mathrm{P}(y_0) = 0,\ \mathrm{P}(x_0) = 0\}.$$

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 $\mathrm{P}(x) = 0$ or $\mathrm{P}(y) = 0$; and hence, $\mathrm{P}(x, y) = 0$. 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.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de