Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 95 (2014)

Small value parallel repetition for general games

Mark Braverman * Ankit Garg †

April 3, 2015

Abstract

We prove a parallel repetition theorem for general games with value tending to 0. Previously Dinur and Steurer proved such a theorem for the special case of projection games. We use information theoretic techniques in our proof. Our proofs also extend to the high value regime (value close to 1) and provide alternate proofs for the parallel repetition theorems of Holenstein and Rao for general and projection games respectively. We also extend the example of Feige and Verbitsky to show that the small-value parallel repetition bound we obtain is tight. Our techniques are elementary in that we only need to employ basic information theory and discrete probability in the small-value parallel repetition proof.

1 Introduction

Parallel repetition theorem is one of the cornerstones of complexity theory. It studies hardness amplification of 2-prover 1-round games. In a 2-prover 1-round game $\mathcal{G}$, there are 2 provers, Alice and Bob, and a verifier. The verifier samples a challenge (x, y) from a joint distribution and gives x to Alice and y to Bob. Alice and Bob answer based on x and y, $(a(x), b(y))$, respectively, and they win the game if some predicate of x, y, a, b is satisfied. The central notion of study is the value of game $\mathrm{val}(\mathcal{G})$, which is the maximum probability of winning over all strategies of Alice and Bob. A natural question is what is the value of n independent parallel repetitions of the game, in other words, is it true that $\mathrm{val}(\mathcal{G}^n) \leq \mathrm{val}(\mathcal{G})^n$? The main difficulty in proving such a theorem arises from the ability of the players to correlate their answers across different coordinates. The first bound on $\mathrm{val}(\mathcal{G}^n)$ was proven by Verbitsky [Ver94] who showed that the value must go to zero as n goes to infinity. Later, Raz [Raz98] proved exponential convergence to zero with the convergence rate depending on the answer length of the game. Feige and Verbitsky [FV02] provided an example to show that the dependence on answer length is necessary. Raz's proof was subsequently simplified and improved by Holenstein [Hol07]. Rao [Rao08] improved Holenstein's proof for the special class of projection games. The techniques of Raz, Holenstein and Rao were information theoretic. Parallel repetition theorem is very useful for gap amplification of PCPs. Rao's theorem for projection games was useful for reducing the Unique Games Conjecture (UGC) to a weaker version.

*Department of Computer Science, Princeton University, email: mbraverm@cs.princeton.edu. Research supported in part by an an NSF CAREER award (CCF-1149888), a Turing Centenary Fellowship, a Packard Fellowship in Science and Engineering, and the Simons Collaboration on Algorithms and Geometry.

†Department of Computer Science, Princeton University, email: garg@cs.princeton.edu. Research supported by a Simons Fellowship in Theoretical Computer Science.

ISSN 1433-8092

Parallel repetition for small value: The proofs of Raz, Holenstein and Rao worked only when the value of the game is close to 1. It wasn't known if a version of parallel repetition could be true when $\mathrm{val}(\mathcal{G})$ is $o(1)$. Dinur and Steurer [DS14] recently proved such a theorem for the special case of projection games, introducing linear-algebraic techniques for parallel repetition along the way. In this paper, we give a proof for a tight parallel repetition theorem in the general small-value case using information theoretic techniques. In the process, we also give an alternative proof for the asymptotically tight bound in the small value projection case, albeit with weaker constants than [DS14].

1.1 Proof overview, intuition, and discussion

We start with a somewhat informal proof outline[1]. Here we opt to gloss over some technical details to convey the main ideas of the proof. This brief exposition is followed by a brief technical overview of the innovations in this proof compared to previous attempts, aimed at those familiar with the previous line of work on parallel repetition. We hope that this exposition will help elucidate our techniques and make them reusable in other related settings.

A high-level overview. All proofs of parallel repetition theorems, including the present one, follow the same high-level strategy: we want to prove that if the value of $\mathcal{G}^n$ is too high, then there is a "too-good-to-be-true" strategy for $\mathcal{G}$. Note that if the optimal strategy $\mathcal{S}_n$ for $\mathcal{G}^n$ were independent over the n coordinates then we would have had $\mathrm{val}(\mathcal{G}^n) = \mathrm{val}(\mathcal{G})^n$, or $\mathrm{val}(\mathcal{G}) = \mathrm{val}(\mathcal{G}^n)^{1/n}$. A more contrived equivalent way of saying this is that if $\mathcal{S}_n$ were independent over the n coordinates, Alice and Bob could have dealt with a challenge (x, y) by embedding (x, y) into a coordinate i of a challenge $((x_1, \ldots, x_n), (y_1, \ldots, y_n))$ (by jointly sampling the remaining pairs (x_{-i}, y_{-i})); having Alice and Bob calculate the strategies $(a_1, \ldots, a_n)$ and $(b_1, \ldots, b_n)$ prescribed by $\mathcal{S}_n$, respectively; and having Alice output a_i and Bob output b_i as their response to the challenge (x, y)). Since $\mathcal{S}_n$ is a product strategy, this clearly works.

The challenge is to make this embedding work even when $\mathcal{S}_n$ is a general strategy where each a_i depends on *the entire vector* $(x_1, \ldots, x_n)$. Note that as we know from counterexamples that it can happen that $\mathrm{val}(\mathcal{G}^n) \gg \mathrm{val}(\mathcal{G})^n$, this is not a mere technicality. Still, while the naïve embedding above breaks down, the general mold of the construction is a valid one: (1) embedding (x, y) into the i-th coordinate for some i; (2) sampling some public information R conditioned on (x, y); (3) having Alice and Bob play according to $\mathcal{S}_n$ conditioned on $(R, x_i = x)$ and $(R, y_i = y)$, respectively; (4) arranging R so that we can prove that the success probability of this strategy is sufficiently high.

Some previous parallel repetition proofs use the assumption that the success probability on coordinate i given success on some other coordinates is high as their departure point, and arrive at a contradiction. By proving that many of these conditional probabilities are low, these proofs establish that the probability of winning all coordinates simultaneously is also low by using the fact that

$$\Pr[\text{win on all coords}] = \prod_{i=1}^{n} \Pr[\text{win on coord } i | \text{win on coords} < i].$$

[1] Note that while we formulate our proofs for the low-value case, as this is the case that had been open, our proof easily extends to match existing proofs for $\mathrm{val}(\mathcal{G})$ close to 1.

Since our main tool is symmetrization, we instead opt to define the random variable 1_W representing whether the players win on *all* n coordinates, and condition everything on the event W that they do win. Under the assumption that $p_W = \Pr[W]$ is not too small, we hope (and eventually prove) that this conditioning does not distort individual coordinates by too much: after sampling the "big" public variable R conditioned on W, Alice and Bob will sample their respective strategies on x_i and y_i ignoring W altogether. This is achieved by a careful choice of what to include in the variable R. This choice is made to balance the parameters of the problem.

The most naïve strategy would have Alice and Bob sample (x_{-i}, y_{-i}) and then play the strategy prescribed by $\mathcal{S}_n$ as they did in the special "product strategy" case above. Unfortunately, this only leads to a success probability of p_W, which is exponentially worse than what we would hope for. We would like to somehow "zoom" on the strategies of Alice and Bob conditioned on W. In other words, they would like to sample (a_i, b_i) conditioned on (x_i, y_i), and W. The problem is that conditioned on W, a_i is very far from being independent from y_i (or, for that matter from b_i conditioned on x_i). This makes such sampling impossible. To address this issue, we will have Alice and Bob sample a public variable R such that conditioned on R and $x_i = x$, a_i is (almost) independent of 1_W and y_i. Thus to sample a_i conditioned on x_i, R and the event W, Alice can ignore W and the fact she doesn't know $y_i = y$, and just sample her strategy conditioned on x_i, R.

The remaining challenge is carefully selecting the variable R. Ignoring W for the moment, we would like a_i conditioned on x_i, R to be independent from y_i. Note that in general the distribution of the answer a_i in $\mathcal{S}_n$ depends on the distribution of all coordinates x_{-i} and not just on $x_i = x$. We could have R empty, and thus have Alice and Bob sample x_{-i} and y_{-i} on their own, but since x_j and y_j are not independent, this would lead to a wrong distribution of inputs to $\mathcal{S}_n$, and thus to a wrong distribution of outputs. Another extreme solution would be to have $R = \{x_{-i}, y_{-i}\}$ contain *all* coordinates except for the i-th one. This would solve the dependence problem, but create a new one: conditioned on W, there could be a very high dependence ($\sim \log 1/\Pr[1_W]$ bits of mutual information) between R and e.g. x_i, thus making it impossible for Alice and Bob (who each only have access to either x_i or y_i but not both) to sample R. As an illustration, consider the following example. Let $M = 1/\Pr[W]$ be an integer, and imagine an n-coordinate game where Alice and Bob win if and only if $\sum_{j=1}^{n}(x_j + y_j) = 0 \mod M$. Then sampling (x_{-i}, y_{-i}) correctly conditioned on W requires the knowledge of $x_i + y_i \mod M$, something neither Alice nor Bob possesses. Our solution is similar to previous solutions, although its exact execution is inspired by the latest developments of information complexity techniques, particularly in the context of direct product for communication complexity. R will contain a set x_G of x's and y_H of y's such that each coordinate $j \neq i$ is contained in $G \cup H$. Thus for each such j either x_j or y_j is publicly sampled. Conditioning on R breaks the dependence between the remaining x's and y's, which can then be sampled privately. Still, R "misses" enough coordinates that the mutual information between R and y_i conditioned on W is small, and thus R can be simultaneously jointly sampled by Alice and Bob (at least with high enough probability).

Such dependence breaking appeared in previous parallel repetition proofs, as well as in information complexity/communication complexity contexts [BYJKS04, BR11]. Here, however, the existence of the arbitrary random variable 1_W on which we are conditioning, creates technical difficulties that do not exists in previous context. We address those by choosing G and H to have a $\Theta(n)$ overlap — a technical innovation that, to the best of our knowledge, was only employed once before [BRWY13], and the potential applications of which are still not fully understood.

An additional complication that we need to address is that even if the mutual information between y_i and R given x_i is "small" (and thus it should be possible for Alice to sample R without knowing y_i), and similarly for Bob, this mutual information will not be very small. In particular, the best we can hope for is something of the form $O((\log 1/\Pr[1_W])/n)$, which, in the small success probability regime, is still $\omega(1)$. In previous (high success probability) works, this mutual information was $o(1)$, and thus the statistical distance between $R|x_i$, $R|y_i$, and the variable $R|(x_i, y_i)$ Alice and Bob really want to sample is also $o(1)$ by Pinsker Inequality. In this case, an approximate sample from $R|(x_i, y_i)$ is obtained using a joint sampling technique employed by Holenstein and Rao [Hol07, Rao08] in their simplified proofs of the parallel repetition theorem. In the low-success-probability case we end up proving the following statement: if the mutual information $I(Y_i; R|X_i)$ and $I(X_i; R|Y_i)$ are $< \log 1/\delta$, then R can be correctly jointly sampled with probability $> poly(\delta)$. Note that this probability is $o(1)$ when $\log 1/\delta = \omega(1)$, but is still high enough for our purposes. More precisely, we sample a distribution that doesn't over-sample any value of R by more than a factor of 2; it is noteworthy that such a sampling is sufficient for our purposes. The sampling lemma we prove may find other applications in complexity theory.

With R having been sampled, Alice and Bob are able to independently sample a_i and b_i conditioned on $(R, x_i = x)$ and $(R, y_i = y)$, respectively. There is one last concern: to win the game, Alice and Bob need to sample (a_i, b_i) conditioned on R, their respective inputs, and the event W. They are only able to sample these conditioned on R and their inputs. Thus, as discussed above, our final goal is to limit the dependence between 1_W and (a_i, b_i). Here we employ a trick that has been used before, though our presentation perhaps shows it in a slightly different light. To reduce the interaction between (a_i, b_i) and 1_W conditioned on R, we "hide" (a_i, b_i) among $\sim T$ other pairs of answers to challenges in $G \cap H$. This reduces the dependence between 1_W and (a_i, b_i) to $O((\log 1/\Pr[1_W])/T)$ bits of information, which becomes small as T increases. However, adding the answers to T creates and additional dependence and adds to $I(Y_i; R|X_i, W)$ and $I(X_i; R|Y_i, W)$. The additional contribution is on the order of $T(\log s)/n$, where s is the size of the answer space (and thus $O(T \log s)$ is the entropy of the publicly sampled answers). Finally, a T is chosen to balance the two constraints.

Discussion of techniques. At a technical level, the present paper further develops the idea of symmetrizing out a dependence through a careful choice of conditioning. Similar to the situation in the study of direct sum and product questions in communication complexity, all we want is to claim that there is a coordinate that is "average" in the effect conditioning on winning has on it. The simplest tool available to us which allows us to make such claims in the information-theoretic domain is the chain rule. Unfortunately, breaking the mutual information of a family of variables using the chain rule produces a family of conditional mutual information expressions, each of which has a different conditioning. The main challenge was thus to select a distribution of conditioning terms consistent with the various chain rules needed in the proof. In particular, as was the case in the proof of the direct product theorem for randomized communication complexity [BRWY13], we seem to need to condition on a family of overlapping variables. Understanding why this is the case, and systematizing the use of such conditioning remains an interesting challenge.

The second technical innovation is a joint sampling procedure for the high information-discrepancy regime. Informally, it allows Alice and Bob who each have a distribution μ_A, μ_B, respectively, such that $D(\mu||\mu_A), D(\mu||\mu_B) \leq k$ to jointly (approximately) sample from μ with probability $> 2^{-O(k)}$. The proof of the lemma is similar to previous low-success probability constructions in [BW12, KLL+12], but its current formulation might be of use elsewhere.

We should note that while the notation is somewhat intimidating, the new proof is completely elementary. It only uses basic probability, repeated applications of the chain rule, and some elementary calculus. In particular, it does not use more advanced tools e.g. from linear algebra or spectral graph theory. Still, it is quite possible to draw parallels between our proof and the algebraic proof of Dinur and Steurer for the projection case [DS14]. This raises the tantalizing possibility of finding deeper connections between spectral and information-theoretic tools, and exploiting tools from one to advance the other.

One challenge involving the parallel repetition theorem the present paper does not address is the gap that is present in the case of general games with value close to 1. Assuming $\mathrm{val}(\mathcal{G}) = 1-\varepsilon$, the best upper bound on $\mathrm{val}(\mathcal{G}^n)$ is $(1-\varepsilon^3)^{\Omega(n/\log s)}$ [Hol07], while the best counterexample [Raz11] only gives a lower bound of $(1-\varepsilon^2)^{O(n)}$. If indeed the lower bound is the tight one (in terms of dependence on ε), it would be interesting to see whether our techniques can be used to prove it.

2 Preliminaries

2.1 Notation

We will use capital letters, e.g. A, B, X, Y to denote random variables. If X is a random variable, we will use P_X to denote its distribution. We will frequently use expectations of mutual information, so we will have a compact notation for it. Suppose $A_1, \ldots, A_n$, $B_1, \ldots, B_n$ and $C_1, \ldots, C_n$ are random variables. Let S, G, H be random subsets of $[n]$. Then we will use the notation:

$$\mathbb{E}_{P_{S,G,H}} I(A_S; B_G | C_H) := \mathbb{E}_{s,g,h \sim P_{S,G,H}} I(A_s; B_g | C_h)$$

Here A_s denotes $(A_i)_{i \in s}$. Also we will use the notation:

$$\mathbb{E}_{P_{C,D}} D(P_{A|C} || P_{B|D}) := \mathbb{E}_{c,d \sim P_{C,D}} D(P_{A|C=c} || P_{B|D=d})$$

2.2 Information theory

In this section we briefly provide the essential information-theoretic concepts required to understand the rest of the paper. For a thorough introduction to the area of information theory, the reader should consult the classical book by Cover and Thomas [CT91]. Unless stated otherwise, all log's in this paper are base-2.

Definition 2.1. Let μ be a probability distribution on sample space Ω. *Shannon entropy* (or just *entropy*) of μ, denoted by $H(\mu)$, is defined as $H(\mu) := \sum_{\omega \in \Omega} \mu(\omega) \log \frac{1}{\mu(\omega)}$.

For a random variable A we shall write $H(A)$ to denote the entropy of the induced distribution on the range of A. The same also holds for other information-theoretic quantities appearing later in this section.

Definition 2.2. *Conditional entropy* of a random variable A conditioned on B is defined as

$$H(A|B) = \mathbb{E}_b(H(A|B = b)).$$

Fact 2.3. $H(AB) = H(A) + H(B|A)$.

Definition 2.4. The *mutual information* between two random variable A and B, denoted by $I(A;B)$ is defined as

$$I(A;B) := H(A) - H(A|B) = H(B) - H(B|A).$$

The *conditional mutual information* between A and B given C, denoted by $I(A;B|C)$, is defined as

$$I(A;B|C) := H(A|C) - H(A|BC) = H(B|C) - H(B|AC).$$

Fact 2.5 (Chain Rule). *Let A_1, A_2, B, C be random variables. Then*

$$I(A_1A_2;B|C) = I(A_1;B|C) + I(A_2;B|A_1C).$$

Definition 2.6. Given two probability distributions μ_1 and μ_2 on the same sample space Ω such that $(\forall\omega \in \Omega)(\mu_2(\omega) = 0 \Rightarrow \mu_1(\omega) = 0)$, the *Kullback-Leibler Divergence* (also known as relative entropy) between them is defined as

$$D(\mu_1||\mu_2) = \sum_{\omega\in\Omega} \mu_1(\omega) \log \frac{\mu_1(\omega)}{\mu_2(\omega)}.$$

Fact 2.7 (Chain Rule for relative entropy). *Let P_{V_1,V_2} and P_{U_1,U_2} be two bivariate distributions. Then*

$$D\left(P_{V_1,V_2}||P_{U_1,U_2}\right) = D\left(P_{V_1}||P_{U_1}\right) + \mathbb{E}_{v_1\sim P_{V_1}} D\left(P_{V_2|V_1=v_1}||P_{U_2|U_1=v_1}\right)$$

Fact 2.8 (Convexity of relative entropy). *Let P_1, P_2, Q_1, Q_2 be distributions and $\lambda \in [0,1]$ be a number. Then*

$$D\left(\lambda P_1 + (1-\lambda)P_2||\lambda Q_1 + (1-\lambda)Q_2\right) \leq \lambda D\left(P_1||Q_1\right) + (1-\lambda)D\left(P_2||Q_2\right)$$

The connection between the mutual information and the Kullback-Leibler divergence is provided by the following fact.

Fact 2.9. *For random variables $A, B,$ and C we have*

$$I(A;B|C) = \mathbb{E}_{b,c}(D(A_{bc}||A_c)).$$

where A_{bc} denotes the random variable $A|B=b, C=c$.

Fact 2.10 (Pinsker's inequality). *Let P, Q be two distributions. Then*

$$D(P||Q) \geq \frac{||P-Q||_1^2}{2\ln 2}$$

Here $||P-Q||_1$ is the l_1 distance between the distributions P and Q.

2.3 Basic facts and lemmas

Fact 2.11. *Let A, B, C, D be random variables s.t. $I(A; D|C) = 0$. Then*

$$I(A; B|C) \leq I(A; B|C, D)$$

Proof. Consider $I(A; B, D|C)$ and expand it using chain rule in two ways. $\square$

The following lemma is well known and is used in parallel repetition proofs of Holenstein and Rao as well. We provide a proof for completeness.

Lemma 2.12. *Let $P_{V_1,\ldots,V_n}$ and $P_{U_1,\ldots,U_n}$ be two distributions over some space $\mathcal{U}^n$. Also suppose that $P_{U_1,\ldots,U_n}$ is a product distribution i.e. $P_{U_1,\ldots,U_n}(u_1,\ldots,u_n) = P_{U_1}(u_1)\cdots P_{U_n}(u_n)$. Then*

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

Proof. By the chain rule for relative entropy, we get that:

$$\begin{aligned}
D\left(P_{V_1,\ldots,V_n}||P_{U_1,\ldots,U_n}\right) &= \sum_{i=1}^{n} \mathbb{E}_{v_1,\ldots,v_{i-1}\sim P_{V_1,\ldots,V_{i-1}}} D\left(P_{V_i|V_1=v_1,\ldots,V_{i-1}=v_{i-1}}||P_{U_i|U_1=v_1,\ldots,U_{i-1}=v_{i-1}}\right) \\
&= \sum_{i=1}^{n} \mathbb{E}_{v_1,\ldots,v_{i-1}\sim P_{V_1,\ldots,V_{i-1}}} D\left(P_{V_i|V_1=v_1,\ldots,V_{i-1}=v_{i-1}}||P_{U_i}\right) \\
&\geq \sum_{i=1}^{n} D\left(P_{V_i}||P_{U_i}\right)
\end{aligned}$$

The second equality is because $P_{U_1,\ldots,U_n}$ is a product distribution. The inequality follows by convexity of relative entropy. $\square$

Fact 2.13. *Let P_U be the distribution of some random variable U and let W be an arbitrary event. Then*

$$D\left(P_{U|W}||P_U\right) \leq \log(1/\Pr[W])$$

Proof.

$$\begin{aligned}
D\left(P_{U|W}||P_U\right) &= \sum_{u} P_{U|W}(u)\log(P_{U|W}(u)/P_U(u)) \\
&\leq \sum_{u} P_{U|W}(u)\log(1/\Pr[W]) = \log(1/\Pr[W])
\end{aligned}$$

$\square$

The following lemma is taken from [BRWY13].

Fact 2.14 ([BRWY13], Lemma 19)**.** *Suppose A, B, C are random variables s.t. $I(A; B|C) = 0$ and W be an arbitrary event. Then*

$$I(A; B|C, W) \leq \log(1/\Pr[W])$$

The following lemma is used in a lot of information complexity papers.

Lemma 2.15. *Let P and Q be distributions over a universe $\mathcal{U}$. Let $\mathcal{B} = \{u : \frac{P(u)}{Q(u)} \geq 2^t\}$. Then*

$$P(\mathcal{B}) \leq \frac{D(P||Q) + 1}{t}$$

Proof.

$$\begin{aligned} D(P||Q) &= \sum_{u \in \mathcal{B}} P(u) \cdot \log(P(u)/Q(u)) + \sum_{u \notin \mathcal{B}} P(u) \cdot \log(P(u)/Q(u)) \\ &\geq P(\mathcal{B}) \cdot t + \sum_{u \notin \mathcal{B}} P(u) \cdot \log(P(u)/Q(u)) \end{aligned} \tag{1}$$

Denote the complement of $\mathcal{B}$ by $\bar{\mathcal{B}}$. Then

$$\begin{aligned} \sum_{u \notin \mathcal{B}} P(u) \cdot \log(P(u)/Q(u)) &\geq P(\bar{\mathcal{B}}) \log(P(\bar{\mathcal{B}})/Q(\bar{\mathcal{B}})) \\ &\geq P(\bar{\mathcal{B}}) \log(P(\bar{\mathcal{B}})) \\ &> -1 \end{aligned} \tag{2}$$

The first inequality follows from log-sum inequality. The second inequality is true because $Q(\bar{\mathcal{B}}) \leq 1$. The third inequality follows from the fact that $x \log(x) > -1$ for all $x \geq 0$. Now combining equations (1) and (2) completes the proof of the lemma. ◻

Fact 2.16. *Let P and Q be distributions over a universe $\mathcal{U}$. Suppose $\mathcal{V} \subseteq \mathcal{U}$ is such that $P(\mathcal{V}) = 1$. Then $Q(\mathcal{V}) \geq 2^{-D(P||Q)}$.*

Proof. It directly follows from the log-sum inequality. Denote the complement of $\mathcal{V}$ by $\bar{\mathcal{V}}$.

$$\begin{aligned} D(P||Q) = \sum_{u \in \mathcal{U}} P(u) \cdot \log(P(u)/Q(u)) &\geq P(\mathcal{V}) \cdot \log(P(\mathcal{V})/Q(\mathcal{V})) + P(\bar{\mathcal{V}}) \cdot \log(P(\bar{\mathcal{V}})/Q(\bar{\mathcal{V}})) \\ &= \log(1/Q(\mathcal{V})) \end{aligned}$$

◻

2.4 Games

Here we formally define a 2-player 1-round game. Such a game $\mathcal{G}$ consists of a verifier and two provers Alice and Bob. The verifier draws (x, y) from some distribution μ on $\mathcal{X} \times \mathcal{Y}$, and gives x to Alice and y to Bob. Alice and Bob answer $a \in \mathcal{A}$ and $b \in \mathcal{B}$ depending on x and y i.e. there exists functions $f : \mathcal{X} \to \mathcal{A}$ and $g : \mathcal{Y} \to \mathcal{B}$ s.t. $a = f(x)$ and $b = g(y)$. They win the game if some predicate of x, y, a, b is satisfied i.e. there exists a subset $V \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{A} \times \mathcal{B}$ such that they win the game if $(x, y, a, b) \in V$. Here V and μ are part of the definition of the game $\mathcal{G}$, so that $\mathcal{G} = (\mathcal{X}, \mathcal{Y}, \mathcal{A}, \mathcal{B}, V, \mu)$. Value of the game $\mathrm{val}(\mathcal{G})$ is defined as the maximum probability of winning over all strategies of Alice and Bob. Formally

$$\mathrm{val}(\mathcal{G}) = \max_{f,g} \Pr_{(x,y)\sim\mu}[(x, y, f(x), g(y)) \in V]$$

The game $\mathcal{G}^n$ is defined as follows: Alice gets $x_1, \ldots, x_n$ and Bob gets $y_1, \ldots, y_n$, where $(x_1, y_1), \ldots, (x_n, y_n)$ are distributed according to μ^n (n independent copies of μ). Alice outputs $a_1, \ldots, a_n = F(x_1, \ldots, x_n)$, where $F : \mathcal{X}^n \to \mathcal{A}^n$ and Bob outputs $b_1, \ldots, b_n = G(y_1, \ldots, y_n)$, where $G : \mathcal{Y}^n \to \mathcal{B}^n$. They win the game if for all i, $(x_i, y_i, a_i, b_i) \in V$. The value is defined similarly:

$$\mathrm{val}(\mathcal{G}^n) = \max_{F,G} \Pr_{(x_1,y_1),\ldots,(x_n,y_n)\sim\mu^n} \left[\bigwedge_{i=1}^{n} ((x_i, y_i, F(x_1, \ldots, x_n)_i, G(y_1, \ldots, y_n)_i) \in V) \right]$$

It is not hard to see that allowing shared randomness between Alice and Bob doesn't change the value of the game. But we'll allow Alice and Bob to use shared randomness to facilitate the proofs. We'll denote the size of the answer set of the game, $|\mathcal{A}| \cdot |\mathcal{B}|$ by s.

There are two special cases of games which are interesting: unique and projection games. A game is unique if its accepting predicate has the following property: for each x, y, a, there exists a unique b s.t. $(x, y, a, b) \in V$. Also for each x, y, b, there exists a unique a s.t. $(x, y, a, b) \in V$. A game is called a projection game if for each x, y, a, there exists a unique b s.t. $(x, y, a, b) \in V$. Note that in a projection game, there might exist multiple accepting answers of Alice corresponding to an answer of Bob, once we fix the questions.

2.5 Previous work

Exponential decay in the value of the game was first proven by Raz [Raz98]. He proved the following theorem:

Theorem 2.17 ([Raz98])**.** *Let $\mathcal{G}$ be a game with $\mathrm{val}(\mathcal{G}) = 1 - \epsilon$ and let $\log(s)$ be the answer size of the game. Then $\mathrm{val}(\mathcal{G}^n) \leq (1 - \epsilon^{32}/2)^{\Omega(n/\log(s))}$.*

This was improved by Holenstein [Hol07] who proved the following theorem:

Theorem 2.18 ([Hol07])**.** *Let $\mathcal{G}$ be a game with $\mathrm{val}(\mathcal{G}) = 1 - \epsilon$ and let $\log(s)$ be the answer size of the game. Then $\mathrm{val}(\mathcal{G}^n) \leq (1 - \epsilon^{3}/2)^{\Omega(n/\log(s))}$.*

Holenstein also proved parallel repetition for no-signaling strategies. Later Rao [Rao08] improved the bound for projection games.

Theorem 2.19 ([Rao08])**.** *Let $\mathcal{G}$ be a projection game with $\mathrm{val}(\mathcal{G}) = 1 - \epsilon$. Then $\mathrm{val}(\mathcal{G}^n) \leq (1 - \epsilon^{2}/2)^{\Omega(n)}$.*

Recently Dinur and Steurer proved parallel repetition for projection games in the small value regime.

Theorem 2.20 ([DS14])**.** *Let $\mathcal{G}$ be a projection game with $\mathrm{val}(\mathcal{G}) = \beta$. Then $\mathrm{val}(\mathcal{G}^n) \leq (4\beta)^{n/4}$.*

There has been a substantial amount of other work on improved parallel repetition for special classes of games, e.g. for free games [BRR$^+$09], expanding games [RR12] and projection games with low threshold rank [TWZ14]. Derandomizing parallel repetition theorems is an important question and there has been some work on it e.g. [Sha13], [DM11]. Recently Moshkovitz [Mos14] has given an operation on projection games, called "fortification", which makes the value of the game to behave nicely under parallel repetition. This enables improvements in the state of the art projection PCP theorem, while bypassing some of the difficulty with general parallel repetition. There also has been a lot of work around parallel repetition for games with entanglement [CSUU08, KV11, DSV14, JPY14, CS14].

3 Results

The main theorem of the paper is the following:

Theorem 3.1. *Let $\mathcal{G}$ be a 2-prover 1-round game. Let s be the size of answer set of the game. If $\mathrm{val}(\mathcal{G}) = \beta$, where $1/s \leq \beta$. Then $\mathrm{val}(\mathcal{G}^n) \leq \beta^{\Omega(n \log(1/\beta)/\log(s))}$, where β is sufficiently small and n sufficiently large.*

The theorem is stated formally in theorem 4.12 below.

Remark 3.2. We assume in the theorem that $\beta \geq 1/s$. Note that this is a very natural assumption, since if for all x, y, there exist a, b s.t. the provers win on x, y, a, b, then provers can just output random answers and they win w.p. $\geq 1/s$. Even without the assumption, a simple reduction can be used to handle the case $\beta < 1/s$. In this case, the bound of the theorem is too strong, as the best we can hope for is a bound of the form $\beta^{\Omega(n)}$. Let $\mathcal{G}_w$ be the sub-game of $\mathcal{G}$ over question pairs (x, y) for which there exists some pair of answers that wins the game. Also let p be the probability that we draw such an (x, y) from the distribution for the game, i.e. p is the probability that game is winnable. Then $\mathrm{val}(\mathcal{G}) = p \cdot \mathrm{val}(\mathcal{G}_w)$ and $\mathrm{val}(\mathcal{G}^n) = p^n \cdot \mathrm{val}(\mathcal{G}_w^n)$. Then if $\mathrm{val}(\mathcal{G}) = \beta$ and $\mathrm{val}(\mathcal{G}_w) = \alpha$, where $\beta < 1/s$. There are two cases: (1) If $\log(1/\alpha) < \log(s)/2 \leq \log(1/\beta)/2$, then $\mathrm{val}(\mathcal{G}) \leq p^n = \beta^n/\alpha^n \leq \beta^{n/2}$. (2) If $\log(1/\alpha) \geq \log(s)/2$, then we can apply theorem 3.1 to the game $\mathcal{G}_w$:

$$\mathrm{val}(\mathcal{G}^n) \leq p^n \cdot \alpha^{c \cdot n \log(1/\alpha)/\log(s)} = p^n \cdot \alpha^{\Omega(n)} \leq (p\alpha)^{\Omega(n)} = \beta^{\Omega(n)}$$

Remark 3.3. $\mathrm{val}(\mathcal{G}^n) \leq \beta^{\Omega(n/\log(1/\beta) \cdot \log(s))}$ is what we'll get if we apply the parallel repetition theorem of Raz [Raz98]. It is not clear how to get $\mathrm{val}(\mathcal{G}^n) \leq \beta^{\Omega(n/\log(s))}$, however even in this bound, there is no "small-value behavior", since $\beta^{\Omega(n/\log(s))} \geq 2^{-\Theta(n)}$, if $\beta \geq 1/s$. However our bound has the "small-value behavior" and it says that we get strong parallel repetition up to constants, if β and $1/s$ are polynomially related.

We also show that Feige and Verbitsky's example [FV02] with tweaking of the parameters proves tightness of theorem 3.1.

Theorem 3.4. *There is a family of games $\mathcal{G}_k$ parametrized by k with $\mathrm{val}(\mathcal{G}_k) = \beta_k \to 0$ s.t. $\mathrm{val}(\mathcal{G}_k^n) \geq \beta_k^{O(n \log(1/\beta_k)/\log(s_k))}$, where $\log(s_k)$ is the answer size of the game $\mathcal{G}_k$ with $\frac{\log(1/\beta_k)}{\log(s_k)} \to 0$.*

Remark 3.5. Theorem 3.1 is clearly tight when $\log(1/\beta) = \Theta(\log(s))$. However we give an example where it is tight even when $\log(1/\beta) = o(\log(s))$.

Remark 3.6. Feige and Verbitsky's example is not tight for games with constant value (it has a slack of $\log\log(s)$). Our work shows that for games with sub-constant value, it is exactly tight upto constant factors.

We also reprove Dinur and Steurer's parallel repetition theorem for projection games in the small value regime. However they get much better constants in their proof. Our proof also extends to the high value regime and it provides an alternate proof for the theorems of Holenstein and Rao.

4 Proof for general games

We will denote by $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ inputs to Alice and Bob respectively in the n copy game. If f, g is a strategy for the game, then we'll denote by $A_1, \ldots, A_n = f(X_1, \ldots, X_n)$ and $B_1, \ldots, B_n = g(Y_1, \ldots, Y_n)$ the answers of Alice and Bob respectively. Let W be the event that they win the game on all coordinates and let 1_W be the indicator random variable for it.

Let S, G, H be random subsets of $[n]$ distributed as follows: Let s_h and s_g be random numbers from $\{3n/4+1, \ldots, n\}$. Let $\sigma : [n] \to [n]$ be a uniformly random permutation. Set $H = \sigma([s_h])$, $G = \sigma(\{n-s_g+1, \ldots, n\})$. Let I be a uniformly random element of $G \cap H$. Let l be a random number from $[T]$, where $T < n/2$ is a parameter. Let S be a uniformly random subset of $G \cap H \backslash \{I\}$ of size l. Let $R_{S,G,H,I}$ denote the random variable $X_{G\backslash\{I\}} Y_{H\backslash\{I\}} A_S B_S$. We will use s, g, h, i to denote instantiations of the random variables S, G, H, I respectively.

Lemma 4.1. $\mathbb{E}_{P_{S,G,H,I}} I(A_I B_I; 1_W | X_I, Y_I, R_{S,G,H,I}) \leq H(1_W)/T$.

Proof. Let $|g \cap h| = m$, and let $l_1, l_2, \ldots, l_m$ be the elements of $g \cap h$. Then the distribution $P_{S,G,H,I}$ can also be described as follows: G, H be distributed as in $P_{S,G,H,I}$. Let κ be a random permutation such that $\kappa(\{l_1, \ldots, l_m\}) = \{l_1, \ldots, l_m\}$, and $t \in_R [T]$. Set $I = \kappa(l_t)$ and $S = \kappa(\{l_{t+1}, \ldots, l_{T+1}\})$. Then

$$
\begin{aligned}
&\mathbb{E}_{P_{S,G,H,I}} I(A_I B_I; 1_W | X_I Y_I R_{S,G,H,I}) \\
&= \mathbb{E}_{P_{G,H}} \mathbb{E}_{\kappa} \mathbb{E}_{t \in_R [T]} I(A_{\kappa(l_t)} B_{\kappa(l_t)}; 1_W | A_{\kappa(\{l_{t+1}, \ldots, l_{T+1}\})} B_{\kappa(\{l_{t+1}, \ldots, l_{T+1}\})} X_G Y_H) \\
&= \mathbb{E}_{P_{G,H}} \mathbb{E}_{\kappa} \frac{1}{T} \sum_{t=1}^{T} I(A_{\kappa(l_t)} B_{\kappa(l_t)}; 1_W | A_{\kappa(\{l_{t+1}, \ldots, l_{T+1}\})} B_{\kappa(\{l_{t+1}, \ldots, l_{T+1}\})} X_G Y_H) \\
&= \frac{1}{T} \mathbb{E}_{P_{G,H}} \mathbb{E}_{\kappa} I(A_{\kappa(\{l_1, \ldots, l_T\})} B_{\kappa(\{l_1, \ldots, l_T\})}; 1_W | A_{\kappa(l_{T+1})} B_{\kappa(l_{T+1})} X_G Y_H) \\
&\leq \frac{H(1_W)}{T}
\end{aligned}
$$

□

Remark 4.2. The variable size of the set S (or the variable sizes of the sets G and H, as we will see in the next lemma) is very important for the symmetrization trick to work (it enables the chain rule via an alternate description of the distribution).

Lemma 4.3. $\mathbb{E}_{P_{S,G,H,I}} I(R_{S,G,H,I}; X_I | Y_I, W) \leq \frac{4}{n} H(1_W)/\Pr[W] + \frac{2(T+1)}{n} \cdot \log(s)$.

Proof. Note that $R_{S,G,H,I}$ consists of two parts : $X_{G\backslash\{I\}} Y_{H\backslash\{I\}}$ and $A_S B_S$. We will prove

$$\mathbb{E}_{P_{S,G,H,I}} I(X_{G\backslash\{I\}} Y_{H\backslash\{I\}}; X_I | Y_I, W) \leq \frac{4}{n} H(1_W)/\Pr[W] \tag{3}$$

and

$$\mathbb{E}_{P_{S,G,H,I}} I(A_S B_S; X_I | X_{G\backslash\{I\}} Y_{H\backslash\{I\}} Y_I, W) \leq \frac{2(T+1)}{n} \cdot \log(s) \tag{4}$$

which together will prove the lemma. To prove the first statement, we first prove the following statement:

$$\mathbb{E}_{P_{G,H,I}} I(X_I; 1_W | Y_I, X_{G\setminus\{I\}} Y_{H\setminus\{I\}}) \leq 4H(1_W)/n$$

The distribution $P_{G,H,I}$ can be seen in the following way: let H be distributed as in $P_{G,H,I}$. Let κ_H be a random permutation that maps $[|H|]$ to H. Choose a random number $l \in \{1, \ldots, n/4\}$. Set $I = \kappa_H(l)$ and $G = \kappa_H(\{l, \ldots, n\})$. Then

$$\begin{aligned}
\mathbb{E}_{P_{G,H,I}} I(X_I; 1_W | Y_I, X_{G\setminus\{I\}} Y_{H\setminus\{I\}}) &= \mathbb{E}_H \mathbb{E}_{\kappa_H} \mathbb{E}_{l \in_R [n/4]} I(X_{\kappa_H(l)}; 1_W | X_{\kappa_H(\{l+1,\ldots,n\})} Y_H) \\
&= \mathbb{E}_H \mathbb{E}_{\kappa_H} \frac{4}{n} \sum_{l=1}^{n/4} I(X_{\kappa_H(l)}; 1_W | X_{\kappa_H(\{l+1,\ldots,n\})} Y_H) \\
&= \frac{4}{n} \mathbb{E}_H \mathbb{E}_{\kappa_H} I(X_{\kappa_H(\{1,\ldots,n/4\})}; 1_W | X_{\kappa_H(\{n/4+1,\ldots,n\})} Y_H) \\
&\leq 4H(1_W)/n
\end{aligned}$$

Now we relate $I(X_i; 1_W | Y_i, X_{g\setminus\{i\}} Y_{h\setminus\{i\}})$ to $I(X_{g\setminus\{i\}} Y_{h\setminus\{i\}}; X_i | Y_i, W)$. Consider $I(X_i; X_{g\setminus\{i\}} Y_{h\setminus\{i\}} 1_W | Y_i)$.

$$\begin{aligned}
&I(X_i; X_{g\setminus\{i\}} Y_{h\setminus\{i\}} 1_W | Y_i) \\
&= I(X_i; X_{g\setminus\{i\}} Y_{h\setminus\{i\}} | Y_i) + I(X_i; 1_W | Y_i X_{g\setminus\{i\}} Y_{h\setminus\{i\}}) \\
&= I(X_i; 1_W | Y_i X_{g\setminus\{i\}} Y_{h\setminus\{i\}}) \qquad (5)
\end{aligned}$$

Also writing it in another way, we get

$$\begin{aligned}
&I(X_i; X_{g\setminus\{i\}} Y_{h\setminus\{i\}} 1_W | Y_i) \\
&= I(X_i; 1_W | Y_i) + I(X_i; X_{g\setminus\{i\}} Y_{h\setminus\{i\}} | Y_i 1_W) \\
&\geq \Pr[W] \cdot I(X_i; X_{g\setminus\{i\}} Y_{h\setminus\{i\}} | Y_i, W) \qquad (6)
\end{aligned}$$

Combining (5) and (6), we get $\mathbb{E}_{P_{S,G,H,I}} I(X_I; X_{G\setminus\{I\}} Y_{H\setminus\{I\}} | Y_I, W) \leq \frac{4}{n} H(1_W)/\Pr[W]$.
To prove (4), notice that the distribution $P_{S,G,H,I}$ can also be described as follows: Let S, H be distributed as in $P_{S,G,H,I}$. Let $\kappa_{S,H}$ be a random permutation conditioned on $\kappa_{S,H}([|S|]) = S$ and $\kappa_{S,H}([|H|]) = H$. Choose a random number l from $\{|S|+1, \ldots, |S|+n/4\}$. Set $I = \kappa_{S,H}(l)$ and $G = S \cup \kappa_{S,H}(\{l, \ldots, n\})$. Then

$$\begin{aligned}
&\mathbb{E}_{P_{S,G,H,I}} I(A_S B_S; X_I | X_{G\setminus\{I\}} Y_{H\setminus\{I\}} Y_I, W) \\
&= \mathbb{E}_{S,H} \mathbb{E}_{\kappa_{S,H}} \mathbb{E}_{l \in_R \{|S|+1,\ldots,|S|+n/4\}} I(A_S B_S; X_{\kappa_{S,H}(l)} | X_{\kappa_{S,H}(\{l+1,\ldots,n\})} X_S Y_H, W) \\
&= \mathbb{E}_{S,H} \mathbb{E}_{\kappa_{S,H}} \frac{4}{n} \sum_{l=|S|+1}^{|S|+n/4} I(A_S B_S; X_{\kappa_{S,H}(l)} | X_{\kappa_{S,H}(\{l+1,\ldots,n\})} X_S Y_H, W) \\
&= \frac{4}{n} \mathbb{E}_{S,H} \mathbb{E}_{\kappa_{S,H}} I(A_S B_S; X_{\kappa_{S,H}(\{|S|+1,\ldots,|S|+n/4\})} | X_{\kappa_{S,H}(\{|S|+n/4+1,\ldots,n\})} X_S Y_H, W) \\
&\leq \frac{4}{n} \mathbb{E}_S H(A_S B_S | W) \\
&\leq \frac{4}{n} \mathbb{E}_S |S| \cdot \log(s) \\
&= \frac{2(T+1)}{n} \cdot \log(s)
\end{aligned}$$

□

Remark 4.4. Note that the similar statement $\mathbb{E}_{P_{S,G,H,I}} I(R_{S,G,H,I}; Y_I | X_I, W) \leq \frac{4}{n} H(1_W)/\Pr[W] + \frac{2(T+1)}{n} \cdot \log(s)$ is also true since the distribution of G and H is symmetric.

The next lemma considers the following situation: Alice and Bob have rough estimates of a distribution P and they want to jointly sample from it. This is very similar to the settings in [BW12] and [KLL+12].

Lemma 4.5. *Suppose Alice knows a distribution P_1 and Bob knows a distribution P_2, and they want to jointly sample from a distribution P (all three are distributions over $\mathcal{U}$). Also $D(P||P_1) \leq \log(1/\eta)$ and $D(P||P_2) \leq \log(1/\eta)$, where $\eta \leq \frac{1}{2}$. Then there is a sampling procedure (using shared randomness) such that*

1. *Suppose Alice outputs p_1 and Bob outputs p_2. There is an event E (which depends just on the shared randomness of the sampling procedure) with $\Pr[E] \geq \eta^{10}$, such that $\Pr[p_1 = p_2 | E] = 1$.*
2. *The distribution of $p_1|E$ is multiplicatively bounded by P i.e. $\forall u$, $\Pr[p_1 = u|E] \leq 2 \cdot P(u)$.*

Proof. Consider the sampling procedure described in protocol 1. Let $\mathcal{A} = \{i \text{ s.t. } q_i < P_1(u_i)/\eta^8\}$, $\mathcal{B} = \{i \text{ s.t. } q_i < P_2(u_i)/\eta^8\}$ and $\mathcal{C} = \{i \text{ s.t. } q_i < P(u_i)\}$. Let E be the event: first index in $\mathcal{A} \cup \mathcal{B}$ lies in $\mathcal{A} \cap \mathcal{B} \cap \mathcal{C}$. Let us first prove that $\Pr[E] \geq \eta^{10}$. Let (u, q) be a uniformly random element of $\mathcal{U} \times [0, 1]$. Then

$$\begin{aligned}
\Pr[E] &= \frac{\Pr[q \leq \min(P_1(u)/\eta^8, P_2(u)/\eta^8, P(u))]}{\Pr[q \leq \max(P_1(u)/\eta^8, P_2(u)/\eta^8)]} \\
&\geq \frac{\Pr[q \leq \min(P_1(u)/\eta^8, P_2(u)/\eta^8, P(u))]}{\Pr[q \leq P_1(u)/\eta^8] + \Pr[q \leq P_2(u)/\eta^8]} \\
&\geq \frac{1}{2} \cdot \eta^8 \cdot |\mathcal{U}| \cdot \Pr[q \leq \min(P_1(u)/\eta^8, P_2(u)/\eta^8, P(u))]
\end{aligned}$$

Let $\mathcal{U}' = \{u \in \mathcal{U} | P(u) \leq \min(P_1(u)/\eta^8, P_2(u)/\eta^8)\}$. Then

$$\Pr[q \leq \min(P_1(u)/\eta^8, P_2(u)/\eta^8, P(u))] \geq \frac{1}{|\mathcal{U}|} \cdot P(\mathcal{U}')$$

and hence

$$\Pr[E] \geq \frac{1}{2} \cdot \eta^8 \cdot P(\mathcal{U}')$$

Since $\Pr_{x \sim P}[P(x)/P_1(x) \geq 1/\eta^8] \leq (\log(1/\eta) + 1)/(8 \log(1/\eta)) \leq 1/4$ (by lemma 2.15), and $\Pr_{x \sim P}[P(x)/P_2(x) \geq 1/\eta^8] \leq 1/4$, we have that $P(\mathcal{U}') \geq 1/2$. Thus $\Pr[E] \geq \eta^{10}$.

1. Using shared randomness, get many uniformly random samples from $\mathcal{U} \times [0, 1]$. Denote these samples by $(u_i, q_i)_{i=1}^{\infty}$.
2. Alice outputs the first u_i s.t. $q_i < P_1(u_i)/\eta^8$ and Bob outputs the first u_j s.t. $q_j < P_2(u_j)/\eta^8$.

Protocol 1: Sampling strategy

Since a subevent of E is the event that first index in $\mathcal{A} \cup \mathcal{B}$ lies in $\mathcal{A} \cap \mathcal{B}$, $\Pr[p_1 = p_2|E] = 1$. It remains to prove $\forall u$, $\Pr[p_1 = u|E] \le 2 \cdot P(u)$.

$$\begin{aligned}
\Pr[p_1 = u|E] &= \frac{\min(P_1(u)/\eta^8, P_2(u)/\eta^8, P(u))}{\sum_{u \in \mathcal{U}} \min(P_1(u)/\eta^8, P_2(u)/\eta^8, P(u))} \\
&\le \frac{P(u)}{\sum_{u \in \mathcal{U}'} \min(P_1(u)/\eta^8, P_2(u)/\eta^8, P(u))} \\
&= \frac{P(u)}{P(\mathcal{U}')} \le 2 \cdot P(u)
\end{aligned}$$

This completes the proof of the lemma. $\square$

Lemma 4.6. $\mathbb{E}_{P_I} D\left(P_{X_I, Y_I|W} || P_{X_I Y_I}\right) \le \frac{\log(1/\Pr[W])}{n}$.

Proof.

$$\begin{aligned}
\mathbb{E}_{P_I} D\left(P_{X_I, Y_I|W} || P_{X_I Y_I}\right) &= \frac{1}{n} \sum_{i=1}^{n} D\left(P_{X_i, Y_i|W} || P_{X_i Y_i}\right) \\
&\le \frac{1}{n} D\left(P_{X_1, Y_1, \ldots X_n, Y_n|W} || P_{X_1, Y_1, \ldots X_n, Y_n}\right) \\
&\le \frac{\log(1/\Pr[W])}{n}
\end{aligned}$$

The first equality is true because P_I is uniform over $[n]$. First inequality follows from lemma 2.12. The second inequality follows from the fact 2.13. $\square$

Lemma 4.7. *Suppose* $2^{-20} \ge \Pr[W] \ge \delta^{n \log(1/\delta)/\log(s)}$*, where* $\delta \ge 1/s^{1/4}$*, and* $n \ge \frac{4\log(s)}{\log(1/\delta)}$*. Fix the parameter* T *in the definition of* $P_{S,G,H,I}$ *to be* $\frac{n \log(1/\delta)}{2\log(s)} - 1$ *(we needed* $T < n/2$ *which is true). Then there exists a fixing of* s, g, h, i *such that:*

1. $\mathbb{E}_{x,y \sim \mu_i} D\left(P_{R_{s,g,h,i}|X_i = x, Y_i = y, W} || P_{R_{s,g,h,i}|X_i = x, W}\right) \le 10 \log(1/\delta)$.
2. $\mathbb{E}_{x,y \sim \mu_i} D\left(P_{R_{s,g,h,i}|X_i = x, Y_i = y, W} || P_{R_{s,g,h,i}|Y_i = y, W}\right) \le 10 \log(1/\delta)$.
3. $D\left(P_{X_i Y_i|W} || P_{X_i Y_i}\right) \le 10 \log(1/\delta)$.
4. $\mathbb{E}_{P_{R_{s,g,h,i}, X_i, Y_i|W}} D\left(P_{A_i B_i|X_i, Y_i, R_{s,g,h,i}, W} || P_{A_i B_i|X_i, Y_i, R_{s,g,h,i}}\right) \le 10 \log(1/\delta)$.

Here μ_i *denotes the distribution* $P_{X_i, Y_i|W}$.

Proof. Lemma 4.3 proves that

$$\mathbb{E}_{P_{S,G,H,I}} I(R_{S,G,H,I}; X_I | Y_I, W) \le \frac{4}{n} H(1_W)/\Pr[W] + \frac{2(T+1)}{n} \cdot \log(s)$$

Similarly one can prove that

$$\mathbb{E}_{P_{S,G,H,I}} I(R_{S,G,H,I}; Y_I | X_I, W) \le \frac{4}{n} H(1_W)/\Pr[W] + \frac{2(T+1)}{n} \cdot \log(s)$$

Since $\Pr[W] \le 2^{-20}$, we have

$$\begin{aligned} H(1_W) &= \Pr[W]\log(1/\Pr[W]) + (1-\Pr[W])\log(1/(1-\Pr[W])) \\ &\le \Pr[W]\log(1/\Pr[W]) + \log(1+2\cdot\Pr[W]) \\ &\le \Pr[W]\log(1/\Pr[W]) + 4\cdot\Pr[W] \\ &\le 1.2\cdot\Pr[W]\log(1/\Pr[W]) \end{aligned}$$

The first inequality follows from $\frac{1}{1-x} \le 1+2x$, for all $0 \le x \le 1/2$. The second inequality is true since $\log(1+2x) \le 4x$, for all $x \ge 0$. The third inequality follows from $\Pr[W] \le 2^{-20}$.

Now we have $T = \frac{n\log(1/\delta)}{2\log(s)} - 1$ and $\frac{2(T+1)}{n}\cdot\log(s) = \log(1/\delta)$. Also

$$\begin{aligned} \frac{4}{n}H(1_W)/\Pr[W] &\le \frac{4\cdot 1.2\cdot\log(1/\Pr[W])}{n} \\ &\le \frac{4\cdot 1.2\cdot\log(1/\delta)^2}{\log(s)} \\ &\le 1.2\log(1/\delta) \end{aligned}$$

Thus

$$\mathbb{E}_{s,g,h,i\sim P_{S,G,H,I}}\mathbb{E}_{x,y\sim\mu_i} D\left(P_{R_{s,g,h,i}|X_i=x,Y_i=y,W}||P_{R_{s,g,h,i}|X_i=x,W}\right) = \\ \mathbb{E}_{P_{S,G,H,I}} I(R_{S,G,H,I};Y_I|X_I,W) \le 2.2\log(1/\delta) \tag{7}$$

Similarly

$$\mathbb{E}_{s,g,h,i\sim P_{S,G,H,I}}\mathbb{E}_{x,y\sim\mu_i} D\left(P_{R_{s,g,h,i}|X_i=x,Y_i=y,W}||P_{R_{s,g,h,i}|Y_i=y,W}\right) \le 2.2\log(1/\delta) \tag{8}$$

By lemma 4.6, we get that

$$\mathbb{E}_{i\sim P_I} D\left(P_{X_iY_i|W}||P_{X_iY_i}\right) \le \log(1/\delta)^2/\log(s) \le \log(1/\delta)/4 \tag{9}$$

Also, by lemma 4.1

$$\mathbb{E}_{P_{S,G,H,I}} I(A_IB_I;1_W|X_I,Y_I,R_{S,G,H,I}) \le H(1_W)/T \tag{10}$$

Note that

$$I(A_iB_i;1_W|X_i,Y_i,R_{s,g,h,i}) \ge \\ \mathbb{E}_{P_{X_i,Y_i,R_{s,g,h,i}}} \Pr[W|X_i,Y_i,R_{s,g,h,i}]\cdot D\left(P_{A_iB_i|X_i,Y_i,R_{s,g,h,i},W}||P_{A_iB_i|X_i,Y_i,R_{s,g,h,i}}\right) \tag{11}$$

Combining (10) and (11), we get

$$\frac{H(1_W)}{T\cdot\Pr[W]} \ge \\ \mathbb{E}_{P_{S,G,H,I}}\mathbb{E}_{P_{X_I,Y_I,R_{S,G,H,I}}}\left(\Pr[W|X_I,Y_I,R_{S,G,H,I}]/\Pr[W]\right)\cdot D\left(P_{A_IB_I|X_I,Y_I,R_{S,G,H,I},W}||P_{A_IB_I|X_I,Y_I,R_{S,G,H,I}}\right) = \\ \mathbb{E}_{P_{S,G,H,I}}\mathbb{E}_{P_{X_I,Y_I,R_{S,G,H,I}|W}} D\left(P_{A_IB_I|X_I,Y_I,R_{S,G,H,I},W}||P_{A_IB_I|X_I,Y_I,R_{S,G,H,I}}\right)$$

Now

$$
\begin{aligned}
\frac{H(1_W)}{T \cdot \Pr[W]} &\leq 1.2 \cdot \frac{\log(1/\Pr[W])}{T} \\
&\leq 1.2 \cdot \frac{n \log(1/\delta)^2}{\log(s)} \cdot \frac{1}{\frac{n\log(1/\delta)}{2\log(s)} - 1} \\
&= 2.4 \cdot \frac{\log(1/\delta)}{1 - \frac{2\log(s)}{n\log(1/\delta)}} \\
&\leq 4.8 \log(1/\delta) \qquad (\text{since } n \geq \frac{4\log(s)}{\log(1/\delta)})
\end{aligned}
$$

This gives:

$$
\mathbb{E}_{s,g,h,i \sim P_{S,G,H,I}} \mathbb{E}_{P_{R_{s,g,h,i},X_i,Y_i|W}} D\left(P_{A_iB_i|X_i,Y_i,R_{s,g,h,i},W} || P_{A_iB_i|X_i,Y_i,R_{s,g,h,i}}\right) \leq 4.8 \log(1/\delta) \tag{12}
$$

Applying a Markov argument to (7), (8), (9) and (12) completes the proof.

□

Lemma 4.8. *Let i satisfy the condition in lemma 4.7 i.e. $D(\mu_i||\mu) \leq 10\log(1/\delta)$, where μ_i is the distribution $P_{X_i,Y_i|W}$ and μ is the distribution P_{X_i,Y_i}. Also suppose $\delta^{120} \leq 1/2$. Then there exists a distribution ν_i s.t. $\nu_i \leq 2 \cdot \mu_i$ and $\mu \geq \delta^{380} \cdot \nu_i$.*

Proof. Let $\mathcal{B} = \{(x,y) | \mu_i(x,y) \geq \mu(x,y)/\delta^{260}\}$. Then $\mu_i(\mathcal{B}) \leq (10\log(1/\delta)+1)/(260\log(1/\delta)) \leq 1/2$ (by lemma 2.15 and $\delta^{120} \leq 1/2$). Now define the distribution ν_i as follows:

$$
\nu_i(x,y) = \begin{cases} 0 & : (x,y) \in \mathcal{B} \\ \frac{\mu_i(x,y)}{1-\mu_i(\mathcal{B})} & : (x,y) \notin \mathcal{B} \end{cases}
$$

It is clear from the definition of ν_i that $\nu_i \leq 2 \cdot \mu_i$. Now, if $(x,y) \in \mathcal{B}$, then clearly $\mu(x,y) \geq \nu_i(x,y) = 0$. If $(x,y) \notin \mathcal{B}$, then $\nu_i(x,y) \leq 2 \cdot \mu_i(x,y) \leq 2 \cdot \frac{1}{\delta^{260}} \cdot \mu(x,y) \leq \mu(x,y)/\delta^{380}$. This completes the proof. □

The next lemma is about breaking dependencies between Alice and Bob, which will be very crucial in the proof of main theorem.

Lemma 4.9. *Let $\mathcal{G}$ be a 2-prover 1-round game. Suppose $(X_1, Y_1), .., (X_n, Y_n)$ are inputs for $\mathcal{G}^n$ and let f, g be a strategy for $\mathcal{G}^n$. Let $A_1, \ldots, A_n = f(X_1, \ldots, X_n)$ and $B_1, \ldots, B_n = g(Y_1, \ldots, Y_n)$. Suppose $G, H, S_a, S_b \subset [n]$ and $i \in [n]$ be such that $G \cup H = [n] \backslash \{i\}$. Then*

$$
P_{A_i,B_i|X_G=\bar{x},Y_H=\bar{y},A_{S_a}=\bar{a},B_{S_b}=\bar{b},X_i=x,Y_i=y} =
$$
$$
P_{A_i|X_G=\bar{x},Y_H=\bar{y},A_{S_a}=\bar{a},X_i=x} \otimes P_{B_i|X_G=\bar{x},Y_H=\bar{y},B_{S_b}=\bar{b},Y_i=y}
$$

if $\Pr[X_G = \bar{x}, Y_H = \bar{y}, A_{S_a} = \bar{a}, B_{S_b} = \bar{b}, X_i = x, Y_i = y] > 0$.

Proof. Note that

$$P_{A_i,B_i|X_G=\bar{x},Y_H=\bar{y},A_{S_a}=\bar{a},B_{S_b}=\bar{b},X_i=x,Y_i=y}(a,b) = P_{A_i|X_G=\bar{x},Y_H=\bar{y},A_{S_a}=\bar{a},B_{S_b}=\bar{b},X_i=x,Y_i=y}(a) \cdot P_{B_i|X_G=\bar{x},Y_H=\bar{y},A_{S_a}=\bar{a},B_{S_b}=\bar{b},X_i=x,Y_i=y,A_i=a}(b)$$

Lets first prove

$$P_{A_i|X_G=\bar{x},Y_H=\bar{y},A_{S_a}=\bar{a},B_{S_b}=\bar{b},X_i=x,Y_i=y}(a) = P_{A_i|X_G=\bar{x},Y_H=\bar{y},A_{S_a}=\bar{a},X_i=x}(a)$$

The other part,

$$P_{B_i|X_G=\bar{x},Y_H=\bar{y},A_{S_a}=\bar{a},B_{S_b}=\bar{b},X_i=x,Y_i=y,A_i=a}(b) = P_{B_i|X_G=\bar{x},Y_H=\bar{y},B_{S_b}=\bar{b},Y_i=y}(b)$$

would follow similarly with the set S_a changed to $S_a \cup \{i\}$.

Let $\mathcal{X}^a$ be the set of $x'_1, \ldots, x'_n$ s.t. $f(x'_1, \ldots, x'_n) = a$, $(x'_j)_{j\in G} = \bar{x}$ and $x'_i = x_i$ i.e. set of all completions of $\bar{x}, x_i$ which evaluate to a under the strategy f. Also let Q be the distribution of $X_1, \ldots, X_n$ conditioned on $X_G = \bar{x}, Y_H = \bar{y}, A_{S_a} = \bar{a}, B_{S_b} = \bar{b}, X_i = x, Y_i = y$. This is the same as distribution of $X_1, \ldots, X_n$ conditioned on $X_G = \bar{x}, Y_H = \bar{y}, A_{S_a} = \bar{a}, X_i = x$, since $[n]\backslash(G \cup \{i\}) \subseteq H$. Denote this distribution by Q'. Then

$$\begin{aligned} P_{A_i|X_G=\bar{x},Y_H=\bar{y},A_{S_a}=\bar{a},B_{S_b}=\bar{b},X_i=x,Y_i=y}(a) &= Q(\mathcal{X}^a) \\ &= Q'(\mathcal{X}^a) \\ &= P_{A_i|X_G=\bar{x},Y_H=\bar{y},A_{S_a}=\bar{a},X_i=x}(a) \end{aligned}$$

□

Remark 4.10. A weaker statement is:

$$P_{A_i,B_i|X_G=\bar{x},Y_H=\bar{y},A_{S_a}=\bar{a},B_{S_b}=\bar{b},X_i=x,Y_i=y} = P_{A_i|X_G=\bar{x},Y_H=\bar{y},A_{S_a}=\bar{a},B_{S_b}=\bar{b},X_i=x} \otimes P_{B_i|X_G=\bar{x},Y_H=\bar{y},A_{S_a}=\bar{a},B_{S_b}=\bar{b},Y_i=y}$$

which is all we will need for the proof of lemma 4.11.

Lemma 4.11. *If $2^{-20} \geq \Pr[W] \geq \delta^{n\log(1/\delta)/\log(s)}$, where $\delta \geq 1/s^{1/4}$, $\delta^{120} \leq 1/2$ and $n \geq \frac{4\log(s)}{\log(1/\delta)}$, then there exists a strategy for winning a single game w.p. $> \delta^{2000}$.*

Proof. Consider the strategy described in protocol 2 for a single copy of the game. We prove that if $\Pr[W] \geq \delta^{n\log(1/\delta)/\log(s)}$, then the strategy wins w.p. $\geq \delta^{1940}$. Let $Q(x,y)$ denote the probability of winning when Alice and Bob get x and y, respectively. Note that the probability of winning is $\mathbb{E}_{x,y\sim\mu}Q(x,y)$. By lemma 4.8, there exists a distribution ν_i s.t. $\nu_i \leq 2 \cdot \mu_i$ and $\mu \geq \delta^{380} \cdot \nu_i$. We will prove that $\mathbb{E}_{x,y\sim\nu_i}Q(x,y) \geq \delta^{1560}$, which will imply that $\mathbb{E}_{x,y\sim\mu}Q(x,y) \geq \delta^{1940}$.

Inputs : Alice gets x, Bob get y, $(x,y) \sim \mu$.

1. Let s, g, h, i be as in lemma 4.7.
2. Alice knows the distribution $P_{R_{s,g,h,i}|X_i=x,W}$ and Bob knows the distribution $P_{R_{s,g,h,i}|Y_i=y,W}$. They use the sampling procedure in lemma 4.5 to sample from $P_{R_{s,g,h,i}|X_i=x,Y_i=y,W}$. Suppose Alice samples r_1 and Bob samples r_2.
3. Alice outputs according to the distribution $P_{A_i|X_i=x,R_{s,g,h,i}=r_1}$ and Bob outputs according the distribution $P_{B_i|Y_i=y,R_{s,g,h,i}=r_2}$.

Protocol 2: Strategy for a single game

Lemma 4.7 together with $\nu_i \leq 2 \cdot \mu_i$ implies that (the lemma applies since $\delta \geq 1/s^{1/4}$):

$$\mathbb{E}_{x,y\sim\nu_i} D\left(P_{R_{s,g,h,i}|X_i=x,Y_i=y,W} || P_{R_{s,g,h,i}|X_i=x,W}\right) \leq 20\log(1/\delta)$$
$$\mathbb{E}_{x,y\sim\nu_i} D\left(P_{R_{s,g,h,i}|X_i=x,Y_i=y,W} || P_{R_{s,g,h,i}|Y_i=y,W}\right) \leq 20\log(1/\delta))$$
$$\mathbb{E}_{x,y\sim\nu_i} \mathbb{E}_{P_{R_{s,g,h,i}|X_i=x,Y_i=y|W}} D\left(P_{A_iB_i|X_i=x,Y_i=y,R_{s,g,h,i},W} || P_{A_iB_i|X_i=x,Y_i=y,R_{s,g,h,i}}\right) \leq 20\log(1/\delta))$$

Let $\mathcal{S} \subset \mathcal{X} \times \mathcal{Y}$ be the set of x, y s.t.

$$D\left(P_{R_{s,g,h,i}|X_i=x,Y_i=y,W} || P_{R_{s,g,h,i}|X_i=x,W}\right) \leq 120\log(1/\delta)$$
$$D\left(P_{R_{s,g,h,i}|X_i=x,Y_i=y,W} || P_{R_{s,g,h,i}|Y_i=y,W}\right) \leq 120\log(1/\delta)$$
$$\mathbb{E}_{P_{R_{s,g,h,i}|X_i=x,Y_i=y|W}} D\left(P_{A_iB_i|X_i=x,Y_i=y,R_{s,g,h,i},W} || P_{A_iB_i|X_i=x,Y_i=y,R_{s,g,h,i}}\right) \leq 120\log(1/\delta)$$

Note that $\nu_i(\mathcal{S}) \geq 1/2$. Fix a pair $x, y \in \mathcal{S}$. We will prove that $Q(x,y) \geq \delta^{1440}$, which will imply that $\mathbb{E}_{x,y\sim\nu_i} Q(x,y) \geq \delta^{1560}$ (since $1/2 \geq \delta^{120}$). Applying lemma 4.5 with $\eta = \delta^{120}$ (note that $\eta \leq 1/2$), we get that there exists an event E with $\Pr[E] \geq \delta^{1200}$, $\Pr[r_1 = r_2|E] = 1$, and the distribution of $r_1|E$ is bounded by $2 \cdot P_{R_{s,g,h,i}|X_i=x,Y_i=y,W}$. This implies that:

$$\mathbb{E}_{r\sim r_1|E} D\left(P_{A_iB_i|X_i=x,Y_i=y,R_{s,g,h,i}=r,W} || P_{A_iB_i|X_i=x,Y_i=y,R_{s,g,h,i}=r}\right) \leq 240\log(1/\delta) \tag{13}$$

Let $\mathcal{G}_{x,y} = \{(a,b)|V(x,y,a,b) = 1\}$, that is the set of accepting answers when the questions are x, y. Note that $P_{A_iB_i|X_i=x,Y_i=y,R_{s,g,h,i}=r,W}(\mathcal{G}_{x,y}) = 1$. This implies (by fact 2.16):

$$P_{A_iB_i|X_i=x,Y_i=y,R_{s,g,h,i}=r}(\mathcal{G}_{x,y}) \geq 2^{-D\left(P_{A_iB_i|X_i=x,Y_i=y,R_{s,g,h,i}=r,W} || P_{A_iB_i|X_i=x,Y_i=y,R_{s,g,h,i}=r}\right)}$$

which along with (13) and convexity of the function 2^{-x} implies that:

$$\mathbb{E}_{r\sim r_1|E} P_{A_iB_i|X_i=x,Y_i=y,R_{s,g,h,i}=r}(\mathcal{G}_{x,y}) \geq \delta^{240}$$

Let $Q_E(x,y)$ be the probability of winning conditioned on event E. A very important observation is that:

$$Q_E(x,y) = \mathbb{E}_{r\sim r_1|E} P_{A_iB_i|X_i=x,Y_i=y,R_{s,g,h,i}=r}(\mathcal{G}_{x,y})$$

This is true because $P_{A_iB_i|X_i=x,Y_i=y,R_{s,g,h,i}=r} = P_{A_i|X_i=x,R_{s,g,h,i}=r} \otimes P_{B_i|Y_i=Y,R_{s,g,h,i}=r}$ (lemma 4.9, it applies since $\Pr[X_i = x, Y_i = y, R_{s,g,h,i} = r] > 0$), and $\Pr[r_1 = r_2|E] = 1$. Thus $Q_E(x, y) \geq \delta^{240}$, and hence $Q(x, y) \geq \Pr[E] \cdot Q_E(x, y) \geq \delta^{1440}$. Note that $P_{A_iB_i|X_i=x,Y_i=y,R_{s,g,h,i}=r} = P_{A_i|X_i=x,R_{s,g,h,i}=r} \otimes P_{B_i|Y_i=Y,R_{s,g,h,i}=r}$ is very crucial for us, otherwise the whole proof breaks down. It is crucial to break the dependencies between Alice and Bob and all the weird conditionings were needed so that this property is true.

$\square$

Theorem 4.12. *Let the probability of winning of single game be β, where $\beta \leq 1/2^{20}$ and $\beta \geq 1/s$. Then probability of winning n copies of the game $\leq \beta^{n\log(1/\beta)/(2000)^2\log(s)}$. Here $n \geq \frac{4\log(s)}{\log(1/\delta)}$.*

Proof. Suppose that $\Pr[W] \geq \beta^{n\log(1/\beta)/(2000)^2\log(s)}$. Then apply lemma 4.11 with $\delta = \beta^{1/2000}$. Since $\beta \leq 1/2^{20}$, we get $\delta^{120} \leq 1/2$ and $\Pr[W] \leq \beta \leq 2^{-20}$. Also since $\beta \geq 1/s$, we have $\delta \geq 1/s^{1/4}$. Note that $\beta^{n\log(1/\beta)/(2000)^2\log(s)} = \delta^{n\log(1/\delta)/\log(s)}$. Hence there exists a strategy for winning a single game w.p. $> \delta^{2000} = \beta$, a contradiction. $\square$

5 Projection games

Theorem 5.1. *Suppose $\mathcal{G}$ is a projection game and $\mathrm{val}(\mathcal{G}) \leq \beta$, for β sufficiently small. Then $\mathrm{val}(\mathcal{G}^n) \leq \beta^{\Omega(n)}$.*

We recall the definition of a projection game. A game is called a projection game if for each x, y, a, there exists a unique b s.t. $(x, y, a, b) \in V$ i.e. the provers win on the tuple (x, y, a, b).
We will denote by $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ inputs to Alice and Bob respectively in the n copy game. If f, g is a strategy for the game, then we'll denote by $A_1, \ldots, A_n = f(X_1, \ldots, X_n)$ and $B_1, \ldots, B_n = g(Y_1, \ldots, Y_n)$ the answers of Alice and Bob respectively. Let W be the event that they win the game on all coordinates and let 1_W be the indicator random variable for it.

We will use a slightly different proof strategy. As before, let S, G, H be random subsets of $[n]$ distributed as follows: Let s_h and s_g be random numbers from $\{3n/4+1, \ldots, n\}$. Let $\sigma : [n] \to [n]$ be a uniformly random permutation. Set $H = \sigma([s_h])$, $G = \sigma(\{n - s_g + 1, \ldots, n\})$. Let I be a uniformly random element of $G \cap H$. Let l be a random number from $[T]$, where $T = n/4$. Let S be a uniformly random subset of $G \cap H \backslash \{I\}$ of size l. Let $L_{S,G,H,I}$ denote the random variable $X_{G\backslash\{I\}} Y_{H\backslash\{I\}} B_S$. **The upshot is that we can afford a larger $\boldsymbol{T (= n/4)}$ here, whereas in the general games proof, we could only afford $\boldsymbol{T = \Theta(n\log(1/\beta)/\log(s))}$.**

Lemma 5.2. $\mathbb{E}_{P_{S,G,H,I}} I(A_I; Y_I | X_I, L_{S,G,H,I}, W) \leq 4 \cdot \frac{\log(1/\Pr[W])}{n}$

Proof. As in the proof of lemma 4.3, the distribution $P_{S,G,H,I}$ can also be described as follows: Let S, G be distributed as in $P_{S,G,H,I}$. Let $\kappa_{S,G}$ be a random permutation conditioned on $\kappa_{S,G}([|S|]) = S$ and $\kappa_{S,G}([|G|]) = G$. Choose a random number l from $\{|S|+1, \ldots, |S|+n/4\}$. Set $I = \kappa_{S,G}(l)$ and

$H = S \cup \kappa_{S,G}(\{l, \ldots, n\})$.

$$
\begin{aligned}
&\mathbb{E}_{P_{S,G,H,I}} I(A_I; Y_I | X_I, L_{S,G,H,I}, W) \\
&= \mathbb{E}_{P_{S,G,H,I}} I(A_I; Y_I | X_G, Y_{H\setminus\{I\}}, B_S, W) \\
&\leq \mathbb{E}_{P_{S,G,H,I}} I(X_{[n]\setminus G}; Y_I | X_G, Y_{H\setminus\{I\}}, B_S, W) \\
&= \mathbb{E}_{S,G} \mathbb{E}_{\kappa_{S,G}} \mathbb{E}_{l \in_R \{|S|+1,\ldots,|S|+n/4\}} I(X_{[n]\setminus G}; Y_{\kappa_{S,G}(l)} | Y_{\kappa_{S,G}(\{l+1,\ldots,n\})}, Y_S, X_G, B_S, W) \\
&= \frac{4}{n} \cdot \mathbb{E}_{S,G} \mathbb{E}_{\kappa_{S,G}} \sum_{l=|S|+1}^{|S|+n/4} I(X_{[n]\setminus G}; Y_{\kappa_{S,G}(l)} | Y_{\kappa_{S,G}(\{l+1,\ldots,n\})}, Y_S, X_G, B_S, W) \\
&= \frac{4}{n} \cdot \mathbb{E}_{S,G} \mathbb{E}_{\kappa_{S,G}} I(X_{[n]\setminus G}; Y_{\kappa_{S,G}(\{|S|+1,\ldots,|S|+n/4\})} | Y_{\kappa_{S,G}(\{|S|+n/4+1,\ldots,n\})}, Y_S, X_G, B_S, W) \\
&\leq 4 \cdot \frac{\log(1/\Pr[W])}{n}
\end{aligned}
$$

The first inequality is true since $X_{[n]}$ determines A_i. The last inequality follows from fact 2.14 and that $I(X_{[n]\setminus g}; Y_{\kappa_{s,g}(\{|s|+1,\ldots,|s|+n/4\})} | Y_{\kappa_{s,g}(\{|s|+n/4+1,\ldots,n\})}, Y_s, X_g, B_s) = 0$. This is because $|g| > 3n/4 > T + n/4 \geq |s| + n/4$, therefore $\kappa_{s,g}(\{|s|+1, \ldots, |s|+n/4\}) \subseteq g$ and hence $([n]\setminus g) \subseteq \kappa_{s,g}(\{|s|+n/4+1, \ldots, n\})$. Note that conditioning on B_s creates dependencies between $Y_1, \ldots, Y_n$, however conditioned on $Y_{[n]\setminus g}$, there is no dependency between $X_{[n]\setminus g}$ and other Y_j's. □

Lemma 5.3. $\mathbb{E}_{P_{S,G,H,I}} I(L_{S,G,H,I}; Y_I | X_I, W) \leq 8 \cdot \frac{\log(1/\Pr[W])}{n}$

Proof. $L_{S,G,H,I}$ consists of two parts: $X_{G\setminus\{I\}} Y_{H\setminus\{I\}}$ and B_S. We know from the proof of lemma 4.3 that

$$\mathbb{E}_{P_{S,G,H,I}} I(Y_I; X_{G\setminus\{I\}} Y_{H\setminus\{I\}} | X_I, W) \leq \frac{4}{n} \cdot \log(1/\Pr[W]) \tag{14}$$

So we care about:

$$
\begin{aligned}
\mathbb{E}_{P_{S,G,H,I}} I(B_S; Y_I | X_G, Y_{H\setminus\{I\}}, W) &\leq \mathbb{E}_{P_{S,G,H,I}} I(A_S; Y_I | X_G, Y_{H\setminus I}, W) \\
&\leq \mathbb{E}_{P_{S,G,H,I}} I(X_{[n]\setminus G}; Y_I | X_G, Y_{H\setminus I}, W)
\end{aligned}
\tag{15}
$$

The first inequality is extremely important and this is where **we use the projection property**. The inequality holds because conditioned on W, X_s, Y_s and A_s determine B_s. Note that we use the fact that $s \subseteq (g \cap h)\setminus\{i\}$. The second inequality is true since $X_{[n]}$ determines A_s. Now by an averaging argument similar to the proof of lemma 5.2, we have that:

$$\mathbb{E}_{P_{S,G,H,I}} I(X_{[n]\setminus G}; Y_I | X_G, Y_{H\setminus I}, W) \leq 4 \cdot \frac{\log(1/\Pr[W])}{n} \tag{16}$$

The only difference from the proof of lemma 5.2 is that we will use

$$I(X_{[n]\setminus g}; Y_{\kappa_{s,g}(\{|s|+1,\ldots,|s|+n/4\})} | Y_{\kappa_{s,g}(\{|s|+n/4+1,\ldots,n\})}, Y_s, X_g) = 0$$

instead of

$$I(X_{[n]\setminus g}; Y_{\kappa_{s,g}(\{|s|+1,\ldots,|s|+n/4\})} | Y_{\kappa_{s,g}(\{|s|+n/4+1,\ldots,n\})}, Y_s, X_g, B_s) = 0.$$

Combining equations (14), (15) and (16) proves the lemma. □

Lemma 5.4. $\mathbb{E}_{P_{S,G,H,I}} I(L_{S,G,H,I}; X_I | Y_I, W) \le 8 \cdot \frac{\log(1/\Pr[W])}{n}$

Proof. The proof of lemma 4.3 gives:

$$\mathbb{E}_{P_{S,G,H,I}} I(X_{G\setminus\{I\}} Y_{H\setminus\{I\}}; X_I | Y_I, W) \le 4 \cdot \frac{\log(1/\Pr[W])}{n} \tag{17}$$

Also

$$\begin{aligned}\mathbb{E}_{P_{S,G,H,I}} I(B_S; X_I | X_{G\setminus I}, Y_H, W) &\le \mathbb{E}_{P_{S,G,H,I}} I(Y_{[n]\setminus H}; X_I | X_{G\setminus I}, Y_H, W) \\ &\le 4 \cdot \frac{\log(1/\Pr[W])}{n} \end{aligned} \tag{18}$$

The first inequality holds because $Y_{[n]}$ determines B_s. The second inequality is similar to the proof of lemma 5.3. Combining equations (17) and (18) proves the lemma. □

Lemma 5.5. $\mathbb{E}_{P_{S,G,H,I}} I(B_I; 1_W | X_I, Y_I, L_{S,G,H,I}, A_I) \le H(1_W)/T = \frac{4H(1_W)}{n}$

Proof. Since $I(B_i; X_{[n]\setminus g} | X_i, Y_i, L_{s,g,h,i}, A_i) = 0$, we have by fact 2.11 that:

$$\begin{aligned} I(B_i; 1_W | X_i, Y_i, L_{s,g,h,i}, A_i) &\le I(B_i; 1_W | X_i, Y_i, L_{s,g,h,i}, A_i, X_{[n]\setminus g}) \\ &\le I(B_i; 1_W | X_i, Y_i, L_{s,g,h,i}, X_{[n]\setminus g}) \end{aligned}$$

The second inequality follows from the fact that A_i is a deterministic function of $X_{[n]}$. Also

$$X_i, Y_i, L_{s,g,h,i}, X_{[n]\setminus g} = X_{[n]}, Y_h, B_s$$

Hence

$$\mathbb{E}_{P_{S,G,H,I}} I(B_I; 1_W | X_I, Y_I, L_{S,G,H,I}, A_I) \le \mathbb{E}_{P_{S,G,H,I}} I(B_I; 1_W | X_{[n]}, Y_H, B_S) \tag{19}$$

As in the proof of lemma 4.1, the distribution $P_{S,G,H,I}$ can also be described as follows: G, H be distributed as in $P_{S,G,H,I}$. Let κ be a random permutation such that $\kappa(\{l_1, \dots, l_m\}) = \{l_1, \dots, l_m\}$, and $t \in_R [T]$. Set $I = \kappa(l_t)$ and $S = \kappa(\{l_{t+1}, \dots, l_{T+1}\})$. Here $G \cap H = \{l_1, \dots, l_m\}$. Now

$$\begin{aligned} \mathbb{E}_{P_{S,G,H,I}} I(B_I; 1_W | X_{[n]}, Y_H, B_S) &= \mathbb{E}_{P_{G,H}} \mathbb{E}_\kappa \mathbb{E}_{t\in_R[T]} I(B_{\kappa(l_t)}; 1_W | X_{[n]}, Y_H, B_{\kappa(\{l_{t+1},\dots,l_{T+1}\})}) \\ &= \frac{1}{T} \cdot \mathbb{E}_{P_{G,H}} \mathbb{E}_\kappa \sum_{t=1}^{T} I(B_{\kappa(l_t)}; 1_W | X_{[n]}, Y_H, B_{\kappa(\{l_{t+1},\dots,l_{T+1}\})}) \\ &= \frac{1}{T} \cdot \mathbb{E}_{P_{G,H}} \mathbb{E}_\kappa I(B_{\kappa(\{l_1,\dots,l_T\})}; 1_W | X_{[n]}, Y_H, B_{\kappa(l_{T+1})}) \\ &\le \frac{H(1_W)}{T} \end{aligned} \tag{20}$$

Combining (19) and (20) completes the proof of the lemma. □

Lemma 5.6. *Let $\mathcal{G}$ be a projection game. Suppose f, g is a strategy for $\mathcal{G}^n$ and let W be the event of winning in all coordinates. If $2^{-20} \ge \Pr[W] \ge \delta^n$, then there exists a fixing of s, g, h, i such that:*

1. $\mathbb{E}_{x,y\sim P_{X_i,Y_i|W}} D\left(P_{L_{s,g,h,i}|X_i=x,Y_i=y,W} || P_{L_{s,g,h,i}|Y_i=y,W}\right) \le O(\log(1/\delta))$

2. $\mathbb{E}_{x,y\sim P_{X_i,Y_i|W}} D\left(P_{L_{s,g,h,i}|X_i=x,Y_i=y,W} || P_{L_{s,g,h,i}|X_i=x,W}\right) \leq O(\log(1/\delta))$

3. $D(P_{X_i,Y_i|W} || P_{X_i,Y_i}) \leq O(\log(1/\delta))$

4. $\mathbb{E}_{x,y\sim P_{X_i,Y_i|W}} \mathbb{E}_{r\sim L_{s,g,h,i}|X_i=x,Y_i=y,W} D\left(P_{A_i|X_i=x,Y_i=y,L_{s,g,h,i}=r,W} || P_{A_i|X_i=x,L_{s,g,h,i}=r,W}\right)$
$\leq O(\log(1/\delta))$

5. $\mathbb{E}_{x,y\sim P_{X_i,Y_i|W}} \mathbb{E}_{r\sim L_{s,g,h,i}|X_i=x,Y_i=y,W}$
$D\left(P_{A_i,B_i|X_i=x,Y_i=y,L_{s,g,h,i}=r,W} || P_{A_i|X_i=x,L_{s,g,h,i}=r,W} \otimes P_{B_i|Y_i=y,L_{s,g,h,i}=r}\right) \leq O(\log(1/\delta))$

Proof. The proof is similar to the proof of lemma 4.7. The proof is a Markov bound applied to the expected versions (expectation over $P_{S,G,H,I}$) of the statements. The expected versions of *1* and *2* follow from lemma 5.4 and 5.3 respectively, as in lemma 4.7. The expected version of *3* is lemma 4.6. The expected version of *4* follows from lemma 5.2. For the expected version of *5*, note that:

$$\begin{aligned}
&\mathbb{E}_{s,g,h,i\sim P_{S,G,H,I}} \mathbb{E}_{x,y\sim P_{X_i,Y_i|W}} \mathbb{E}_{r\sim L_{s,g,h,i}|X_i=x,Y_i=y,W} \\
&D\left(P_{A_i,B_i|X_i=x,Y_i=y,L_{s,g,h,i}=r,W} || P_{A_i|X_i=x,L_{s,g,h,i}=r,W} \otimes P_{B_i|Y_i=y,L_{s,g,h,i}=r}\right) \\
&= \mathbb{E}_{s,g,h,i\sim P_{S,G,H,I}} \mathbb{E}_{x,y\sim P_{X_i,Y_i|W}} \mathbb{E}_{r\sim L_{s,g,h,i}|X_i=x,Y_i=y,W} D\left(P_{A_i|X_i=x,Y_i=y,L_{s,g,h,i}=r,W} || P_{A_i|X_i=x,L_{s,g,h,i}=r,W}\right) \\
&+ \mathbb{E}_{s,g,h,i\sim P_{S,G,H,I}} \mathbb{E}_{x,y\sim P_{X_i,Y_i|W}} \mathbb{E}_{r\sim L_{s,g,h,i}|X_i=x,Y_i=y,W} \mathbb{E}_{a\sim P_{A_i|X_i=x,Y_i=y,L_{s,g,h,i}=r,W}} \\
&D\left(P_{B_i|A_i=a,X_i=x,Y_i=y,L_{s,g,h,i}=r,W} || P_{B_i|Y_i=y,L_{s,g,h,i}=r}\right) \\
&\leq O(\log(1/\delta)) + O(\log(1/\delta)) \\
&= O(\log(1/\delta))
\end{aligned}$$

The first inequality is expected version of *4*. The second inequality we prove below, which will complete the proof of the lemma. We want to prove that:

$$\mathbb{E}_{s,g,h,i\sim P_{S,G,H,I}} \mathbb{E}_{P_{X_i,Y_i,L_{s,g,h,i},A_i|W}} D\left(P_{B_i|X_i,Y_i,L_{s,g,h,i},A_i,W} || P_{B_i|Y_i,L_{s,g,h,i}}\right) \leq O(\log(1/\delta))$$

which is the same as

$$\mathbb{E}_{s,g,h,i\sim P_{S,G,H,I}} \mathbb{E}_{P_{X_i,Y_i,L_{s,g,h,i},A_i|W}} D\left(P_{B_i|X_i,Y_i,L_{s,g,h,i},A_i,W} || P_{B_i|X_i,Y_i,L_{s,g,h,i},A_i}\right) \leq O(\log(1/\delta))$$

since by lemma 4.9, $P_{B_i|X_i,Y_i,L_{s,g,h,i},A_i}$ is the same as $P_{B_i|Y_i,L_{s,g,h,i}}$. Now note that:

$$\begin{aligned}
&\frac{4H(1_W)}{n\Pr[W]} \geq \mathbb{E}_{s,g,h,i\sim P_{S,G,H,I}} \frac{I(B_i;1_W|X_i,Y_i,L_{s,g,h,i},A_i)}{\Pr[W]} \\
&\geq \mathbb{E}_{s,g,h,i\sim P_{S,G,H,I}} \mathbb{E}_{P_{X_i,Y_i,L_{s,g,h,i},A_i}} \frac{\Pr[W|X_i,Y_i,L_{s,g,h,i},A_i]}{\Pr[W]} \cdot D\left(P_{B_i|X_i,Y_i,L_{s,g,h,i},A_i,W} || P_{B_i|X_i,Y_i,L_{s,g,h,i},A_i}\right) \\
&= \mathbb{E}_{s,g,h,i\sim P_{S,G,H,I}} \mathbb{E}_{P_{X_i,Y_i,L_{s,g,h,i},A_i|W}} D\left(P_{B_i|X_i,Y_i,L_{s,g,h,i},A_i,W} || P_{B_i|X_i,Y_i,L_{s,g,h,i},A_i}\right) \qquad (21)
\end{aligned}$$

The first inequality is lemma 5.5. The second inequality follows by writing mutual information as an expected divergence. Now since $Pr[W] \leq 2^{-20}$, $\frac{4H(1_W)}{n\Pr[W]} \leq O\left(\frac{\log(1/\Pr[W])}{n}\right) \leq O(\log(1/\delta))$, which completes the proof. □

Lemma 5.7. *Let $\mathcal{G}$ be a projection game. Suppose* $\mathrm{val}(\mathcal{G}^n) \geq \delta^n$*, for δ sufficiently small, then* $\mathrm{val}(\mathcal{G}) \geq \delta^{O(1)}$.

Proof. The proof is very similar to proof of lemma 4.11. We use the strategy for $\mathcal{G}^n$ to obtain a strategy for $\mathcal{G}$. Suppose $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ be inputs to Alice and Bob in $\mathcal{G}^n$ and $A_1, \ldots, A_n$ and $B_1, \ldots, B_n$ be their answers. W be the event of winning on all copies. Consider the strategy defined in protocol 3. Let $Q(x, y)$ denote the probability of winning when Alice and Bob get x and y, respectively. The probability of winning is $\mathbb{E}_{x,y\sim\mu} Q(x, y)$. Let μ_i denote the distribution $P_{X_i,Y_i|W}$. Since by lemma 5.6, $D(\mu_i||\mu) \leq O(\log(1/\delta))$, we get by lemma 4.8, there exists a distribution ν_i s.t. $\nu_i \leq 2 \cdot \mu_i$ and $\mu \geq \delta^{O(1)} \cdot \nu_i$. We'll prove that $\mathbb{E}_{x,y\sim\nu_i} Q(x, y) \geq \delta^{O(1)}$, which will imply that $\mathbb{E}_{x,y\sim\mu} Q(x, y) \geq \delta^{O(1)}$.

Inputs : Alice gets x, Bob get y, $(x, y) \sim \mu$.

1. Let s, g, h, i be as in lemma 5.6.
2. Alice knows the distribution $P_{L_{s,g,h,i}|X_i=x,W}$ and Bob knows the distribution $P_{L_{s,g,h,i}|Y_i=y,W}$. They use the sampling procedure in lemma 4.5 to sample from $P_{L_{s,g,h,i}|X_i=x,Y_i=y,W}$. Suppose Alice samples r_1 and Bob samples r_2.
3. Alice outputs according to the distribution $P_{A_i|X_i=x,L_{s,g,h,i}=r_1,W}$ and Bob outputs according the distribution $P_{B_i|Y_i=y,L_{s,g,h,i}=r_2}$.

Protocol 3: Strategy for a single game: Projection case

Lemma 5.6 together with $\nu_i \leq 2 \cdot \mu_i$ implies that:

$$\mathbb{E}_{x,y\sim\nu_i} D\left(P_{L_{s,g,h,i}|X_i=x,Y_i=y,W} || P_{L_{s,g,h,i}|Y_i=y,W}\right) \leq O(\log(1/\delta))$$
$$\mathbb{E}_{x,y\sim\nu_i} D\left(P_{L_{s,g,h,i}|X_i=x,Y_i=y,W} || P_{L_{s,g,h,i}|X_i=x,W}\right) \leq O(\log(1/\delta))$$
$$\mathbb{E}_{x,y\sim\nu_i} \mathbb{E}_{r\sim L_{s,g,h,i}|X_i=x,Y_i=y,W} D\left(P_{A_i,B_i|X_i=x,Y_i=y,L_{s,g,h,i}=r,W} || P_{A_i|X_i=x,L_{s,g,h,i}=r,W} \otimes P_{B_i|Y_i=y,L_{s,g,h,i}=r}\right)$$
$$\leq O(\log(1/\delta))$$

Let $\mathcal{S} \subset \mathcal{X} \times \mathcal{Y}$ be the set of x, y s.t.

$$D\left(P_{L_{s,g,h,i}|X_i=x,Y_i=y,W} || P_{L_{s,g,h,i}|Y_i=y,W}\right) \leq 6 \cdot O(\log(1/\delta)) = O(\log(1/\delta))$$
$$D\left(P_{L_{s,g,h,i}|X_i=x,Y_i=y,W} || P_{L_{s,g,h,i}|X_i=x,W}\right) \leq 6 \cdot O(\log(1/\delta)) = O(\log(1/\delta))$$
$$\mathbb{E}_{r\sim L_{s,g,h,i}|X_i=x,Y_i=y,W} D\left(P_{A_i,B_i|X_i=x,Y_i=y,L_{s,g,h,i}=r,W} || P_{A_i|X_i=x,L_{s,g,h,i}=r,W} \otimes P_{B_i|Y_i=y,L_{s,g,h,i}=r}\right)$$
$$\leq 6 \cdot O(\log(1/\delta)) = O(\log(1/\delta))$$

Then $\nu_i(\mathcal{S}) \geq 1/2$. Fix a pair $x, y \in \mathcal{S}$. We will prove that $Q(x, y) \geq \delta^{O(1)}$, which will imply that $\mathbb{E}_{x,y\sim\nu_i} Q(x, y) \geq \delta^{O(1)}$, for δ sufficiently small. Applying lemma 4.5 with $\eta = \delta^{O(1)}$ (note that $\eta \leq 1/2$ for δ sufficiently small), we get that there exists an event E with $\Pr[E] \geq \delta^{O(1)}$, $\Pr[r_1 = r_2|E] = 1$, and the distribution of $r_1|E$ is bounded by $2 \cdot P_{L_{s,g,h,i}|X_i=x,Y_i=y,W}$. This implies that:

$$\mathbb{E}_{r\sim r_1|E} D\left(P_{A_i,B_i|X_i=x,Y_i=y,L_{s,g,h,i}=r,W} || P_{A_i|X_i=x,L_{s,g,h,i}=r,W} \otimes P_{B_i|Y_i=y,L_{s,g,h,i}=r}\right) \leq O(\log(1/\delta)) \tag{22}$$

Let $\mathcal{G}_{x,y} = \{(a,b)|V(x,y,a,b)=1\}$, that is the set of accepting answers when the questions are x, y. Note that $P_{A_iB_i|X_i=x,Y_i=y,L_{s,g,h,i}=r,W}(\mathcal{G}_{x,y}) = 1$. This implies (by fact 2.16):

$$\begin{aligned} & P_{A_i|X_i=x,L_{s,g,h,i}=r,W} \otimes P_{B_i|Y_i=y,L_{s,g,h,i}=r}(\mathcal{G}_{x,y}) \\ & \geq 2^{-D\left(P_{A_i,B_i|X_i=x,Y_i=y,L_{s,g,h,i}=r,W} || P_{A_i|X_i=x,L_{s,g,h,i}=r,W} \otimes P_{B_i|Y_i=y,L_{s,g,h,i}=r}\right)} \end{aligned}$$

which along with (22) and convexity of the function 2^{-x} implies that:

$$\mathbb{E}_{r\sim r_1|E} P_{A_i|X_i=x,L_{s,g,h,i}=r,W} \otimes P_{B_i|Y_i=y,L_{s,g,h,i}=r}(\mathcal{G}_{x,y}) \geq \delta^{O(1)} \tag{23}$$

Let $Q_E(x,y)$ be the probability of winning conditioned on event E. Then by (23), $Q_E(x,y) \geq \delta^{O(1)}$, which implies $Q(x,y) \geq \Pr[E] \cdot Q_E(x,y) \geq \delta^{O(1)}$.

□

Proof. (Of theorem 5.1) Follows from lemma 5.7. □

6 Unique games

For unique games, we can obtain a simpler proof for the following theorem:

Theorem 6.1. *Let $\mathcal{G}$ be a unique game. Then if $\mathrm{val}(\mathcal{G}) = \beta$, then for β sufficiently small, $\mathrm{val}(\mathcal{G}^n) \leq \beta^{\Omega(n)}$.*

The idea is the same as in the proof of general games, but we can afford to sample $\Omega(n)$ answers. Let S, G, H be random subsets of $[n]$ distributed as follows: Let s_h and s_g be random numbers from $\{3n/4+1, \ldots, n\}$. Let $\sigma : [n] \to [n]$ be a uniformly random permutation. Set $H = \sigma([s_h])$, $G = \sigma(\{n - s_g + 1, \ldots, n\})$. Let I be a uniformly random element of $G \cap H$. Let l be a random number from $[T]$, where $T < n/2$ is a parameter. Let S be a uniformly random subset of $G \cap H \backslash \{I\}$ of size l. Let $R_{S,G,H,I}$ denote the random variable $X_{G\backslash\{I\}} Y_{H\backslash\{I\}} A_S B_S$. Here we will choose $T = n/4$. Lemma 4.1 gives us:

$$\mathbb{E}_{P_{S,G,H,I}} I(A_I B_I; 1_W | X_I Y_I R_{S,G,H,I}) \leq H(1_W)/T = O(\Pr[W] \cdot \log(1/\beta))$$

The other term we want to analyze is from lemma 4.3: $\mathbb{E}_{P_{S,G,H,I}} I(R_{S,G,H,I}; X_I | Y_I, W)$. Here the analysis slightly deviates from the proof of general games. **We use the following property of unique games**: conditioned on W, X_i, Y_i, A_i fixes B_i and similarly X_i, Y_i, B_i fixes A_i. This is the only place where we will use the unique game property. It affects the analysis of the following term in the proof of lemma 4.3 (rest of proof remains the same).

$$
\begin{aligned}
&\mathbb{E}_{P_{S,G,H,I}} I(A_S B_S; X_I | X_{G\setminus\{I\}} Y_{H\setminus\{I\}} Y_I, W) \\
&= \mathbb{E}_{P_{S,G,H,I}} I(B_S; X_I | X_{G\setminus\{I\}} Y_{H\setminus\{I\}} Y_I, W) + \mathbb{E}_{P_{S,G,H,I}} I(A_S; X_I | X_{G\setminus\{I\}} Y_{H\setminus\{I\}} Y_I, B_S, W) \\
&= \mathbb{E}_{P_{S,G,H,I}} I(B_S; X_I | X_{G\setminus\{I\}} Y_{H\setminus\{I\}} Y_I, W) \\
&\le \mathbb{E}_{P_{S,G,H,I}} I(Y_{[n]\setminus H}; X_I | X_{G\setminus\{I\}} Y_{H\setminus\{I\}} Y_I, W) \\
&\le O\left(\frac{\log(1/\Pr[W])}{n}\right) \\
&\le O(\log(1/\beta))
\end{aligned}
$$

The second inequality follows because $S \subset G\setminus\{I\}$ and $S \subset H\setminus\{I\}$, hence $H(A_S | X_{G\setminus\{I\}} Y_{H\setminus\{I\}} Y_I, B_S, W) = 0$, by the observation about unique games. The rest of the steps are similar to the proofs in the projection games section. This gives us:

$$\mathbb{E}_{P_{S,G,H,I}} I(R_{S,G,H,I}; X_I | Y_I, W) \le O(\log(1/\beta))$$

and

$$\mathbb{E}_{P_{S,G,H,I}} I(R_{S,G,H,I}; Y_I | X_I, W) \le O(\log(1/\beta))$$

from where we can finish the proof similar to the one for general games.

7 Tight lower bound

Theorem 7.1. *There is a family of games $\mathcal{G}_k$ parametrized by k with $\mathrm{val}(\mathcal{G}_k) = \beta_k \to 0$ s.t. $\mathrm{val}(\mathcal{G}_k^n) \ge \beta_k^{O(n\log(1/\beta_k)/\log(s_k))}$, where $\log(s_k)$ is the answer size of the game $\mathcal{G}_k$ with $\frac{\log(1/\beta_k)}{\log(s_k)} \to 0$.*

We show that different parameters in Feige and Verbitsky's counterexample [FV02] give a tight lower matching theorem 3.1. We describe our example below (based on [FV02], we just tweak the parameters):

- There is a parameter k and another parameter $r = k^{1/3}$.
- There is a bipartite graph G where each side has k^r vertices, the properties needed of this bipartite graph will be described later.
- Alice and Bob get uniformly distributed $(x, y) \in_R [k] \times [k]$.
- Alice needs to output $(s_a, l_a) \in [k]^r \times [r]$ and Bob needs to output $(s_b, l_b) \in [k]^r \times [r]$. They win the game if $l_a = l_b$, $s_a(l_a) = x$, $s_b(l_b) = y$ and there is an edge between s_a and s_b in G. The answer length of the game $\log(s) = 2(r\log(k) + \log(r)) = \Theta(k^{1/3}\log(k))$. Lets call this game $\mathcal{G}_k$.
- The properties we need from the graph G are the following: (1) It has at least $k^{2r}/2k^{1/5}$ edges. (2) Every k by k vertex induced subgraph of G has at most $k^2/k^{1/10}$ edges.

We'll prove the existence of such a graph G later. First lets use it to obtain a tight lower bound.

Lemma 7.2. $\mathrm{val}(\mathcal{G}_k^n) \geq \left(\frac{1}{2k^{1/5}}\right)^{n/r}$

Proof. Divide the n copies into chunk of size r each. We'll give a strategy which is independent over different chunks and wins w.p. $\geq \frac{1}{2k^{1/5}}$ in each chunk and this will prove the lemma. Suppose in a chunk Alice gets $\bar{x} = x_1, \ldots, x_r$ and Bob gets $\bar{y} = y_1, \ldots, y_r$. Then Alice outputs $(\bar{x}, 1), \ldots, (\bar{x}, r)$ and Bob outputs $(\bar{y}, 1), \ldots, (\bar{y}, r)$. The players win the all the copies in the chunk if there is an edge between $\bar{x}$ and $\bar{y}$ in G which happens w.p. $\geq \frac{1}{2k^{1/5}}$, since this is the fraction of edges in the graph G. □

Lemma 7.3. $\mathrm{val}(\mathcal{G}_k) \leq 1/k^{1/20}$

Proof. Fix a strategy f, g for $\mathcal{G}_k$. We define a k by k bipartite graph G'. There is an edge between x and y if the players win under the strategy f, g on inputs x and y. Note that $\mathrm{val}(\mathcal{G}_k) = \frac{\#\text{ of edges in } G'}{k^2}$. Suppose $f(x) = (s_a, l_a)$ and $g(y) = (s_b, l_b)$. There is an edge between x and y iff $l_a = l_b$, $s_a(l_a) = x$, $s_b(l_b) = y$ and there is an edge between s_a and s_b in G. Now look at a connected component of G' and the answer (s, l) corresponding to a vertex v in the connected component. l should be the same for all vertices in the component, and also it should hold that $s(l) = v$ for all vertices v. Because of this the answer strings corresponding to vertices on Alice's side in the component are all distinct and similarly for Bob's side. Also # of edges in the component $\leq k^2/k^{1/10}$ because of the property of G. Thus G' has the property that every connected component has at most $k^2/k^{1/10}$ edges. Now using the following claim, we get that $\mathrm{val}(\mathcal{G}_k) = \frac{\#\text{ of edges in } G'}{k^2} \leq 1/k^{1/20}$

Claim 7.4. *Let G' be a k by k bipartite graph with the property that every connected component has at most $\delta \cdot k^2$ edges. Then G' has at most $\sqrt{\delta} \cdot k^2$ edges.*

Proof. (Of claim) Let $c_1, \ldots, c_t$ be the number of vertices in the components. Then $\sum_{i=1}^t c_i = 2k$. In each component, the number of edges $\leq \min\{c_i^2/4, \delta \cdot k^2\}$, since in a bipartite graph with c vertices, number of edges $\leq c^2/4$. Then number of edges in the graph:

$$\begin{aligned} &\leq \sum_{i=1}^{t} \min\{c_i^2/4, \delta \cdot k^2\} \leq \sum_{i=1}^{t} \sqrt{(c_i^2/4) \cdot \delta \cdot k^2} \\ &= \sqrt{\delta} \cdot \left(\sum_{i=1}^{t} c_i\right) \cdot k/2 = \sqrt{\delta} \cdot k^2 \end{aligned}$$

□

□

If $\mathrm{val}(\mathcal{G}_k) = \beta$, lemma 7.2 and 7.3 give

$$\mathrm{val}(\mathcal{G}_k^n) \geq \beta^{\Theta(n/r)} = \beta^{\Theta(n \log(k)/\log(s))} = \beta^{\Theta(n \log(1/\beta)/\log(s))}$$

Now let us prove that a graph G with required properties exists. We want it to have at least $k^{2r}/2k^{1/5}$ edges and every k by k induced subgraph to have at most $k^2/k^{1/10}$ edges. Pick a random graph with each edge included w.p. $1/k^{1/5}$. Then it has at least $k^{2r}/2k^{1/5}$ edges w.p. $1-o(1)$. The probability that some k by k induced subgraph has at least $k^2/k^{1/10}$ edges is:

$$\begin{aligned}
\leq \binom{k^r}{k}^2 \cdot \binom{k^2}{k^2/k^{1/10}} \cdot \left(\frac{1}{k^{1/5}}\right)^{k^{19/10}} &\leq \frac{k^{2rk} \cdot 2^{H(1/k^{1/10}) \cdot k^2}}{2^{k^{19/10}\log(k)/5}} \\
&\leq \frac{k^{2k^{4/3}} \cdot 2^{k^{19/10}\log(k)/8}}{2^{k^{19/10}\log(k)/5}} = o(1)
\end{aligned}$$

The third inequality follows from the fact that for large enough k, $H(1/k^{1/10}) \leq \frac{\log(k)}{8k^{1/10}}$. Since both the bad events occur w.p. $o(1)$, the required graph exists.

8 Games with value close to 1

We provide an alternate proof for the parallel repetition theorem of Holenstein [Hol07].

Theorem 8.1 ([Hol07])**.** *Let $\mathcal{G}$ be a game with $\mathrm{val}(\mathcal{G}) = 1-\epsilon$ and let $\log(s)$ be the answer size of the game. Then $\mathrm{val}(\mathcal{G}^n) \leq (1-\epsilon^3)^{\Omega(n/\log(s))}$, if $n \geq \log(s)/\epsilon^3$ and $\epsilon <= 1/2$.*

The proof techniques for the small value regime readily extend to the case when $\mathrm{val}(\mathcal{G}) = 1-\epsilon$. The only difference is that we have to replace our sampling lemma 4.5 with the correlated sampling lemma of Holenstein [Hol07]. The following variant of the lemma is proven in [Rao08].

Lemma 8.2. *Suppose Alice knows a distribution P_1 and Bob knows a distribution P_2 such that $||P-P_1||_1 \leq \epsilon$ and $||P-P_2||_1 \leq \epsilon$. Then there is a sampling procedure s.t.*

1. *Suppose Alice outputs p_1 and Bob outputs p_2. There exists an event E with $\Pr[E] \geq 1-O(\epsilon)$ s.t. $\Pr[p_1 = p_2|E] = 1$.*
2. *The distribution of $p_1|E$ is P.*

Let us provide a rough sketch of our proof strategy for the high value case. Suppose W be the event of winning in all coordinates. We want to show that $\Pr[W] \leq 2^{-\Omega\left(\epsilon^3 n/\log(s)\right)}$. Assume on the contrary. As in the proof of the small value case, let S, G, H be random subsets of $[n]$ distributed as follows: Let s_h and s_g be random numbers from $\{3n/4+1, \ldots, n\}$. Let $\sigma : [n] \to [n]$ be a uniformly random permutation. Set $H = \sigma([s_h])$, $G = \sigma(\{n-s_g+1, \ldots, n\})$. Let I be a uniformly random element of $G \cap H$. Let l be a random number from $[T]$, where $T < n/2$ is a parameter. Let S be a uniformly random subset of $G \cap H \backslash \{I\}$ of size l. Let $R_{S,G,H,I}$ denote the random variable $X_{G\backslash\{I\}} Y_{H\backslash\{I\}} A_S B_S$. We'll choose $T = \epsilon^2 n/\log(s)$ here.

Recall that the proof of lemma 4.7 gives us:

$$\begin{aligned}
&\mathbb{E}_{P_{S,G,H,I}} \mathbb{E}_{P_{X_I,Y_I,R_{S,G,H,I}|W}} D\left(P_{A_I B_I|X_I,Y_I,R_{S,G,H,I},W} || P_{A_I B_I|X_I,Y_I,R_{S,G,H,I}}\right) \leq \frac{H(1_W)}{T \cdot \Pr[W]} \\
&\leq O(\epsilon) + \frac{1-\Pr[W]}{T \cdot \Pr[W]} \cdot \log\left(\frac{1}{1-\Pr[W]}\right) \\
&\leq O(\epsilon) + O(1/T) \\
&\leq O(\epsilon)
\end{aligned}$$

The last inequality is true for $n \geq \log(s)/\epsilon^3$. Similarly following other steps of lemma 4.7, we will get the following analogue to it: there exists a fixing of s, g, h, i s.t.

$$\mathbb{E}_{x,y\sim\mu_i} D\left(P_{R_{s,g,h,i}|X_i=x,Y_i=y,W}||P_{R_{s,g,h,i}|X_i=x,W}\right) \leq O(\epsilon^2) \tag{24}$$

$$\mathbb{E}_{x,y\sim\mu_i} D\left(P_{R_{s,g,h,i}|X_i=x,Y_i=y,W}||P_{R_{s,g,h,i}|Y_i=y,W}\right) \leq O(\epsilon^2) \tag{25}$$

$$D\left(P_{X_iY_i|W}||P_{X_iY_i}\right) \leq O(\epsilon^2) \tag{26}$$

$$\mathbb{E}_{P_{R_{s,g,h,i},X_i,Y_i|W}} D\left(P_{A_iB_i|X_i,Y_i,R_{s,g,h,i},W}||P_{A_iB_i|X_i,Y_i,R_{s,g,h,i}}\right) \leq O(\epsilon) \tag{27}$$

Here μ_i denotes the distribution $P_{X_i,Y_i|W}$. Then consider the strategy described in protocol 4 for a single copy. We will prove that it wins w.p. $1 - O(\epsilon)$ w.r.t. the distribution μ, which will lead to a contradiction (after scaling ϵ appropriately).

Inputs : Alice gets x, Bob get y, $(x, y) \sim \mu$.

1. Let s, g, h, i be as described above.
2. Alice knows the distribution $P_{R_{s,g,h,i}|X_i=x,W}$ and Bob knows the distribution $P_{R_{s,g,h,i}|Y_i=y,W}$. They use the sampling procedure in lemma 8.2 to sample from $P_{R_{s,g,h,i}|X_i=x,Y_i=y,W}$. Suppose Alice samples r_1 and Bob samples r_2.
3. Alice outputs according to the distribution $P_{A_i|X_i=x,R_{s,g,h,i}=r_1}$ and Bob outputs according the distribution $P_{B_i|Y_i=y,R_{s,g,h,i}=r_2}$.

Protocol 4: Strategy for a single game: high value case

By equation (26) and Pinsker's inequality, we have that: $||\mu_i - \mu||_1 \leq O(\epsilon)$. Thus it is enough to say that the strategy in protocol 4 wins w.p. $1 - O(\epsilon)$ w.r.t. μ_i. Suppose

$$\begin{aligned}
p_{x,y} &:= ||P_{R_{s,g,h,i}|X_i=x,Y_i=y,W} - P_{R_{s,g,h,i}|X_i=x,W}||_1 \\
l_{x,y} &:= ||P_{R_{s,g,h,i}|X_i=x,Y_i=y,W} - P_{R_{s,g,h,i}|Y_i=y,W}||_1 \\
D_{x,y} &:= \mathbb{E}_{P_{R_{s,g,h,i}|X_i=x,Y_i=y,W}} D\left(P_{A_iB_i|X_i=x,Y_i=y,R_{s,g,h,i},W}||P_{A_iB_i|X_i=x,Y_i=y,R_{s,g,h,i}}\right)
\end{aligned}$$

By equation (24), Pinsker's inequality and convexity of the function $f(z) = z^2$, we get $E_{x,y\sim\mu_i} p_{x,y} \leq O(\epsilon)$. Similarly, $E_{x,y\sim\mu_i} l_{x,y} \leq O(\epsilon)$. Also equation (27) gives us that $E_{x,y\sim\mu_i} D_{x,y} \leq O(\epsilon)$. Now fix a particular x, y and look at the probability of winning $Q(x, y)$. The claim is that

$$Q(x, y) \geq 1 - O(l_{x,y} + p_{x,y} + D_{x,y}) \tag{28}$$

This is enough to prove that $\mathbb{E}_{x,y\sim\mu_i} Q(x, y) \geq 1 - O(\epsilon)$, which is what we need. So let us prove (28). By lemma 8.2, there exists an event $E_{x,y}$ s.t. $\Pr[r_1 = r_2|E_{x,y}] = 1$, $r_1|E \sim P_{R_{s,g,h,i}|X_i=x,Y_i=y,W}$ and $\Pr[E_{x,y}] \geq 1 - O(l_{x,y} + p_{x,y})$. Let $Q_E(x, y)$ be the probability of winning conditioned on $E_{x,y}$. By fact 2.16 and convexity of the function $f(z) = 2^{-z}$, we have that:

$$Q_E(x, y) \geq 2^{-D_{x,y}} \geq 1 - O(D_{x,y})$$

Then

$$Q(x,y) \geq \Pr[E_{x,y}] \cdot Q_E(x,y) \geq (1 - O(l_{x,y} + p_{x,y})) \cdot (1 - O(D_{x,y})) \geq 1 - O(l_{x,y} + p_{x,y} + D_{x,y})$$

This completes the proof sketch.

Remark 8.3. The proofs for unique and projection games for the small value case extend similarly to the high value case.

Remark 8.4. A remarkable feature of our proof for the high value case (a property that seems essential in the small value regime) is that we don't need the players to sample $P_{R_{s,g,h,i}|X_i=x,Y_i=y,W}$ conditioned on an event E of probability $1-O(\epsilon)$. It would have sufficed for our purposes to samples from a distribution which is multiplicatively bounded by $P_{R_{s,g,h,i}|X_i=x,Y_i=y,W}$ (say by a factor of 2) conditioned on E. However we don't know yet how to exploit it and it would be interesting if this can lead to improvements for parallel repetition for general and free games in the value-close-to-1 regime. Note that an improved proof has to work around the tightness of the bound for unique and projection games implied by Raz's counterexample [Raz08].

References

[BR11] Mark Braverman and Anup Rao. Information equals amortized communication. In *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on*, pages 748–757. IEEE, 2011.

[BRR+09] Boaz Barak, Anup Rao, Ran Raz, Ricky Rosen, and Ronen Shaltiel. Strong parallel repetition theorem for free projection games. *RANDOM*, 2009.

[BRWY13] Mark Braverman, Anup Rao, Omri Weinstein, and Amir Yehudayoff. Direct products in communication complexity. In *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on*, pages 746–755. IEEE, 2013.

[BW12] Mark Braverman and Omri Weinstein. A discrepancy lower bound for information complexity. In *RANDOM*, pages 459–470. Springer, 2012.

[BYJKS04] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *Journal of Computer and System Sciences*, 68(4):702–732, 2004.

[CS14] André Chailloux and Giannicola Scarpa. Parallel repetition of entangled games with exponential decay via the superposed information cost. *41st International Colloquium on Automata, Languages and Programming*, 2014.

[CSUU08] Richard Cleve, William Slofstra, Falk Unger, and Sarvagya Upadhyay. Perfect parallel repetition theorem for quantum xor proof systems. *Journal of Computational Complexity*, 17(2):282–299, May 2008.

[CT91] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley series in telecommunications. J. Wiley and Sons, New York, 1991.

[DM11] Irit Dinur and Or Meir. Derandomized parallel repetition via structured pcps. *IEEE Conference on Computational Complexity*, 2011.

[DS14] Irit Dinur and David Steurer. Analytical approach to parallel repetition. *46th Annual Symposium on the Theory of Computing*, 2014.

[DSV14] Irit Dinur, David Steurer, and Thomas Vidick. A parallel repetition theorem for entangled projection games. *IEEE Conference on Computational Complexity*, 2014.

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

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

[JPY14] Rahul Jain, Attila Pereszlényi, and Penghui Yao. A parallel repetition theorem for entangled two-player one-round games under product distributions. *IEEE Conference on Computational Complexity*, 2014.

[KLL+12] Iordanis Kerenidis, Sophie Laplante, Virginie Lerays, Jérémie Roland, and David Xiao. Lower bounds on information complexity via zero-communication protocols and applications. In *Foundations of Computer Science (FOCS), 2012 IEEE 53rd Annual Symposium on*, pages 500–509. IEEE, 2012.

[KV11] Julia Kempe and Thomas Vidick. Parallel repetition of entangled games. *43rd annual ACM symposium on Theory of computing*, 2011.

[Mos14] Dana Moshkovitz. Parallel repetition of fortified games. *Electronic Colloquium on Computational Complexity (ECCC)*, 2014.

[Rao08] Anup Rao. Parallel repetition in projection games and a concentration bound. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing*, 2008.

[Raz98] Ran Raz. A parallel repetition theorem. *SIAM Journal on Computing*, 27(3):763–803, June 1998. Prelim version in STOC '95.

[Raz08] Ran Raz. A counterexample to strong parallel repetition. In *Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science*. IEEE Computer Society, 2008.

[Raz11] Ran Raz. A counterexample to strong parallel repetition. *SIAM Journal on Computing*, 40(3):771–777, 2011.

[RR12] Ran Raz and Ricky Rosen. A strong parallel repetition theorem for projection games on expanders. *IEEE Conference on Computational Complexity*, pages 247–257, 2012.

[Sha13] Ronen Shaltiel. Derandomized parallel repetition theorems for free games. *Complexity*, 22(3):565–594, 2013.

[TWZ14] Madhur Tulsiani, John Wright, and Yuan Zhou. Optimal strong parallel repetition for projection games on low threshold rank graphs. *ICALP*, 2014.

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

ECCC ISSN 1433-8092

http://eccc.hpi-web.de