THEORY OF COMPUTING, Volume 5 (2009), pp. 141–172
www.theoryofcomputing.org

Parallel Repetition: Simplifications and the No-Signaling Case*

Thomas Holenstein†

Received: June 20, 2008; published: July 29, 2009.

Abstract: Consider a game where a referee chooses (x,y) according to a publicly known distribution, sends x to Alice, and y to Bob. Without communicating with each other, Alice responds with a value a and Bob responds with a value b. Alice and Bob jointly win if a publicly known predicate $Q(x,y,a,b)$ is satisfied.

Assume that the maximum probability that Alice and Bob can win in such a game is $v < 1$. Raz (SIAM J. Comput. 27, 1998) shows that if the game is repeated n times in parallel, then the probability that Alice and Bob win *all* games simultaneously is at most $\bar{v}^{\frac{n}{\log(s)}}$, where s is the maximal number of possible responses from Alice and Bob in the initial game, and $\bar{v} < 1$ is a constant depending only on v.

In this work, we simplify Raz's proof in various ways and thus shorten it significantly. Further we study the case where Alice and Bob are not restricted to local computations and can use any strategy which does not imply communication between them.

ACM Classification: F.4.1, F.1.2

AMS Classification: 68Q15, 60C05

Key words and phrases: parallel repetition, probabilistically checkable proofs

1 Introduction

In a two-player refereed game, a referee chooses (x,y) randomly according to a known distribution P_{XY}, sends x to Alice, and y to Bob, who respond with values a and b, respectively. Alice and Bob jointly

*A preliminary version of this paper appeared in the proceedings of the thirty-ninth annual ACM Symposium on Theory of Computing (STOC '07) [21]

†This work was done while the author was at ETH Zürich.

 DOI: 10.4086/toc.2009.v005a008

win if a known predicate $Q(x,y,a,b)$ is satisfied; we denote the maximum winning probability for such a game by v and assume $v < 1$.

In the parallel repetition of such a game, n pairs (x_i, y_i) are chosen independently according to $(\mathsf{P}_{XY})^n$. The values $x_1, \ldots, x_n$ are sent to Alice, the values $y_1, \ldots, y_n$ to Bob. The players respond with values $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$, and win if $Q(x_i, y_i, a_i, b_i)$ is satisfied *for all i*.

The question studied in this paper is how much parallel repetition reduces the winning probability of the players. It is motivated by the study of two-prover interactive proofs, initiated by Ben-Or et al. [5]. It was first conjectured that in a game which is repeated n times in parallel, the probability that Alice and Bob win all the games simultaneously is at most v^n (see [19]). However, later a counterexample to this conjecture was given [18].[1]

Related work Because of the PCP-Theorem and its application to hardness of approximation [14, 2, 1], people became interested to know whether the winning probability would at least decrease exponentially in n. Various papers give upper bounds on the winning probability of a game which is repeated n times in parallel [8, 13, 25, 34]; but all of them fall short to give a result of the desired form. Raz then gave a very strong result in [32], showing that the winning probability is at most $\bar{v}^{n/\log(s)}$ for some constant $\bar{v} < 1$ depending only on v, and where s is the total number of answers Alice and Bob may give in the initial game. It is the only explicit bound for arbitrary distributions P_{XY} and quantitatively the strongest. Parnafes, Raz, and Wigderson [29] modify Raz's proof to show that the term $\log(s)$ can be replaced by a parameter which is much smaller for some games. The current paper is based on Raz's result (see below), and gives a slightly stronger upper bound of the same form.

Games for which the n-fold parallel repetition decreases the winning probability less than from v to v^n were also constructed: Fortnow [18] gives a game for which the maximal winning probability in two repetitions is larger than v^2 (see also [16]), Feige [13] constructs a game where the winning probability in two parallel repetitions does not decrease at all, and Feige and Verbitsky [17] give, for infinitely many s, a game where $\Theta(\log(s)/\log\log(s))$ repetitions decrease the winning probability from at most $3/4$ to some value larger than $1/8$, where s is the number of possible answers Alice and Bob can give. This last result shows that in general a dependence on s as in Raz's bound is needed.

No-signaling strategies No-signaling strategies are all those strategies which do not imply communication. Popescu and Rohrlich [30] give an example of such a strategy: Alice receives a bit x, Bob receives a bit y, and they respond with uniform random bits a and b such that $a \oplus b = x \wedge y$. It is clear that this strategy cannot be implemented with shared randomness and without communication. On the other hand, black-box access to the strategy does not give Alice and Bob the power to communicate.

The study of no-signaling strategies is motivated by the idea that if Alice and Bob share some entangled quantum state, the set of possible strategies they might use increases, but stays a subset of the no-signaling strategies (this subset is strict: for example the above strategy which achieves $a \oplus b = x \wedge y$ from (x,y) cannot be simulated perfectly using quantum mechanics [9], see also [28, Problem 2.3]—the corresponding game is called the CHSH-game [10]).

We remark that there are games which can be won with probability 1 given a shared quantum state

[1] For readers not familiar with such counterexamples, a variation of Fortnow's game is reproduced in Section 10.1.

(and thus with a no-signaling strategy), but not using local strategies. Those are called "pseudo-telepathy games" (see [6] and the references therein).

A parallel repetition theorem for the case where Alice and Bob share a quantum state and the decision of the referee only depends on the XOR of the binary answers of Alice and Bob was given by Cleve et al. [11].

Consistent sampling One of the contributions of this paper is to introduce consistent sampling in the context of parallel repetition (Lemma 5.2). It essentially considers the following scenario: assume Alice and Bob are given distributions P_{X_A} and P_{X_B} with the promise that $\|\mathsf{P}_{X_A} - \mathsf{P}_{X_B}\|_1 \leq \varepsilon$, and they are supposed to use shared randomness to produce elements X_A and X_B distributed according to their respective distributions while maximizing $\Pr[X_A = X_B]$; we will see that $1 - O(\varepsilon)$ is achievable.

After the conference version [21] of this paper has been published, it has come to our attention that methods for this are known under the name *consistent sampling*. The first reference seems to be [27], other occurrences are [7, 24, 20, 26]. Often, the special case where the two distributions are uniform over a certain set is studied—the difference to the case considered here is small, however.

Contributions of this paper As mentioned above, the strongest upper bound on the winning probability of a game repeated n times in parallel previous to this paper was given by Raz [32]. His proof was somewhat complicated. In this paper we simplify Raz's proof and give a slightly stronger bound in the following sense: Raz shows that the winning probability of the players is at most $(1 - \Omega((1-v)^c))^{n/\log(s)}$, where $c = 32$ is implicit in the paper (i. e., using our previous notation, Raz shows $\bar{v} = 1 - \Omega((1-v)^{32})$). We improve the constant in the exponent to $c = 3$. The most important change we do to achieve this is to replace a large part (essentially Section 6) of Raz's paper with consistent sampling.

The use of consistent sampling also makes the rest of the argument simpler. We briefly explain why: the main part of Raz's proof consists of showing that the information the players get in the n-fold repetition does not help them win the subgame in some coordinate j, even conditioned on the event that certain other subgames are won. This is done in three steps. In two of these steps the information does not help the players because they can generate this information themselves with local computation only. Consistent sampling implies that this also holds for the third step. This allows us to merge some of the steps, which simplifies the overall structure.

We then study how much the term $\log(s)$ in the exponent in the parallel repetition theorem can be reduced. In [29] it is shown that the logarithm of the partition number of the acceptance predicate can be used instead of $\log(s)$. Based on the ideas from there, we give a bound which might be stronger for some games (see Theorem 8.2).

Finally, we prove a parallel repetition theorem in case Alice and Bob are restricted to no-signaling strategies (in both the given game and the parallel repetition of it).

Subsequent work Theorem 2.5, the main result of this paper, states that the winning probability of the two players in the n-fold repetition is at most $(1-(1-v)^3)^{\Omega(n/s)}$. Based on the proof given here, Rao [31] strengthens this to $(1-(1-v)^2)^{\Omega(n)}$, for games where for every triple (x, y, a) a unique b

satisfies $Q(x,y,a,b)$; such games are known as "projection games." This special case is important in the application of the Parallel Repetition Theorem to probabilistically checkable proofs.

Raz [33] subsequently showed that Rao's bound is essentially tight for the family of odd cycle games, excluding the possibility of reducing the exponent 2 in the above bound. Such an improvement would have been interesting: [15] shows that any exponent smaller than 2 would imply an equivalence between certain Max-Cut hardness and Khot's Unique Games conjecture [23]. Barak et al. [3] strengthen Raz's result, and give a non-trivial approximation algorithm for the value of a unique game (the special case of a projection game where also for each b there is a unique a which satisfies $Q(x,y,a,b)$) repeated in parallel.

In [31], Rao also gives a concentration bound: he shows that it is not only unlikely that the players win *all* the games, but in fact it is unlikely that they win more than a $1-v+\delta$ fraction of the games for any constant $\delta > 0$. For this, he modifies the proof given here appropriately.

2 Notation and basic facts

2.1 Probability distributions

We use calligraphic letters to denote sets. We denote random variables using capital letters, and values with lower case letters. We use superscripts to denote tuples, e. g., $X^n := (X_1,\ldots,X_n)$ and $x^n := (x_1,\ldots,x_n)$.

If a distribution P_{XY} over $\mathcal{X}\times\mathcal{Y}$ is given, we write P_X or P_Y to denote the marginal distribution, e. g., $\mathsf{P}_X(x) := \sum_{y\in\mathcal{Y}}\mathsf{P}_{XY}(x,y)$. The conditional distribution $\mathsf{P}_{Y|X=x}$ is $\mathsf{P}_{Y|X=x}(y) := \mathsf{P}_{XY}(x,y)/\mathsf{P}_X(x)$ which is defined if $\mathsf{P}_X(x) > 0$. To also define it if $\mathsf{P}_X(x) = 0$, we assign expressions of the form $0/0$ the value 0, if they occur in this (or any other) context.

Let P_{X_0} be a distribution over $\mathcal{X}$ and $\mathsf{P}_{Y_1|X_1=x}$ be a conditional distribution over $\mathcal{Y}$. We define the distribution $\mathsf{P}_{X_0}\mathsf{P}_{Y_1|X_1}$ over $\mathcal{X}\times\mathcal{Y}$ as

$$(\mathsf{P}_{X_0}\mathsf{P}_{Y_1|X_1})(x,y) := \mathsf{P}_{X_0}(x)\cdot\mathsf{P}_{Y_1|X_1=x}(y)\,. \tag{2.1}$$

For this, it is necessary that $\mathsf{P}_{Y_1|X_1=x}$ is defined for every $x\in\mathcal{X}$. We also use this notation when $\mathsf{P}_{Y_1|X_1=x}$ is defined as the marginal of a given distribution $\mathsf{P}_{X_1Y_1}$. In this case, we define $\mathsf{P}_{Y_1|X_1=x}$ in an arbitrary way if $\mathsf{P}_{X_1}(x) = 0$. This notation is used for example in Corollary 5.3 in the form $\mathsf{P}_{X_0Y_0}\mathsf{P}_{S|X}$, where it is understood as $(\mathsf{P}_{X_0Y_0}\mathsf{P}_{S|X})(x,y,s) := \mathsf{P}_{X_0Y_0}(x,y)\mathsf{P}_{S|X=x}(s)$. Note that the conditional distribution $\mathsf{P}_{S|X=x}$ is defined there by the marginal distribution P_{SX} of the given distribution P_{SXY}. The notation $\mathsf{P}_{X_0Y_0}\mathsf{P}_{S|X}$ is not completely explicit, since it does not specify that X_0 is to be associated with X. However, it will always be clear from the context.

For two probability distributions P_{X_0} and P_{X_1} over the same set $\mathcal{X}$ we define the statistical distance

$$\|\mathsf{P}_{X_0}-\mathsf{P}_{X_1}\| := \frac{1}{2}\sum_{x\in\mathcal{X}}\big|\mathsf{P}_{X_0}(x)-\mathsf{P}_{X_1}(x)\big|\,. \tag{2.2}$$

The following lemma states alternate ways of characterizing the statistical distance. Equation (2.3) states that the statistical distance is the advantage of the best function in predicting whether a given sample is from X_0 or X_1. Equation (2.4) states that X_0 and X_1 can be generated at the same time such that they differ only with probability $\|\mathsf{P}_{X_0} - \mathsf{P}_{X_1}\|$ (i. e., it provides some sort of coupling).

Lemma 2.1. *Let P_{X_0} and P_{X_1} be distributions over a finite set $\mathcal{X}$. Then,*

$$\max_{f:\mathcal{X}\to\{0,1\}} \big(\Pr[f(X_0)=0] - \Pr[f(X_1)=0]\big) = \|\mathsf{P}_{X_0} - \mathsf{P}_{X_1}\| \tag{2.3}$$

$$\min_{\mathsf{P}_{X_0'X_1'}:\mathsf{P}_{X_0}=\mathsf{P}_{X_0'}\wedge\mathsf{P}_{X_1}=\mathsf{P}_{X_1'}} \big(\Pr[X_0' \neq X_1']\big) = \|\mathsf{P}_{X_0} - \mathsf{P}_{X_1}\|. \tag{2.4}$$

Proof. For (2.3) we refer to [12], equation (11.137). For (2.4) we get, for any $\mathsf{P}_{X_0'X_1'}$ with $(\mathsf{P}_{X_0} = \mathsf{P}_{X_0'}) \wedge (\mathsf{P}_{X_1} = \mathsf{P}_{X_1'})$:

$$\begin{aligned}\Pr[X_0' = X_1'] &= \sum_{x\in\mathcal{X}} \Pr[X_0' = X_1' = x] \le \sum_{x\in\mathcal{X}} \min(\mathsf{P}_{X_0}(x), \mathsf{P}_{X_1}(x)) \\ &= \sum_{x\in\mathcal{X}} \frac{\mathsf{P}_{X_0}(x) + \mathsf{P}_{X_1}(x)}{2} - \frac{|\mathsf{P}_{X_0}(x) - \mathsf{P}_{X_1}(x)|}{2} = 1 - \|\mathsf{P}_{X_0} - \mathsf{P}_{X_1}\|.\end{aligned}$$

We can reach equality. For this, let $\varepsilon := \|\mathsf{P}_{X_0} - \mathsf{P}_{X_1}\|$ and assume $0 < \varepsilon < 1$ (in the other cases the lemma is trivial). Define the probability distributions

$$\begin{aligned}\mathsf{P}_{X} &= \frac{1}{1-\varepsilon}\min(\mathsf{P}_{X_0}, \mathsf{P}_{X_1}), \\ \mathsf{P}_{\widetilde{X}_0} &= \frac{1}{\varepsilon}(\mathsf{P}_{X_0} - \min(\mathsf{P}_{X_0}, \mathsf{P}_{X_1})), \\ \mathsf{P}_{\widetilde{X}_1} &= \frac{1}{\varepsilon}(\mathsf{P}_{X_1} - \min(\mathsf{P}_{X_0}, \mathsf{P}_{X_1})).\end{aligned}$$

Then, $\mathsf{P}_{X_0'X_1'}(x_0, x_1) = (1-\varepsilon)\delta_{x_0,x_1}\mathsf{P}_X(x_0) + \varepsilon\mathsf{P}_{\widetilde{X}_0}(x_0)\mathsf{P}_{\widetilde{X}_1}(x_1)$, where $\delta_{a,b}$ is the Kronecker-delta, has the desired properties (note that it is a convex combination of the two distributions $\delta_{x_0,x_1}P_X(x_0)$ and $\mathsf{P}_{\widetilde{X}_0}(x_0)\mathsf{P}_{\widetilde{X}_1}(x_1)$). □

2.2 Games

Definition 2.2. A *game* $\mathfrak{G} = (\mathsf{P}_{XY}, Q)$ *over* $\mathcal{X}\times\mathcal{Y}\times\mathcal{A}\times\mathcal{B}$ is a distribution P_{XY} over $\mathcal{X}\times\mathcal{Y}$ and a predicate Q over $\mathcal{X}\times\mathcal{Y}\times\mathcal{A}\times\mathcal{B}$. The *value* $v(\mathfrak{G})$ of a game is

$$v(\mathfrak{G}) := \max_{h_a,h_b} \Pr_{XY}[Q(X, Y, h_a(X), h_b(Y))],$$

where the maximization is over functions $h_a : \mathcal{X} \to \mathcal{A}$ and $h_b : \mathcal{Y} \to \mathcal{B}$. A *strategy* (h_a, h_b) for a game is a pair of such functions.

Sometimes also randomized strategies for Alice and Bob are considered, where h_a and h_b also depend on (the same) shared randomness r chosen according to some distribution P_R. However, there always exists an $r \in \mathcal{R}$ such that

$$\begin{aligned}\Pr_{RXY}[Q(X,Y,h_a(X,R),h_b(Y,R))] &= \mathop{\mathrm{E}}_{R}\big[\Pr_{XY}[Q(X,Y,h_a(X,R),h_b(Y,R))]\big] \\ &\le \Pr_{XY}[Q(X,Y,h_a(X,r),h_b(Y,r))], \end{aligned} \tag{2.5}$$

and we see that the definition of the value is robust against such a change. Individual (local) randomness can be obtained from shared randomness and is thus a special case of the above.

Definition 2.3. The *n-fold parallel repetition* $\mathfrak{G}^n$ of a game $\mathfrak{G} = (\mathsf{P}_{XY}, Q)$ over $\mathcal{X} \times \mathcal{Y} \times \mathcal{A} \times \mathcal{B}$ is the game over $\mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{A}^n \times \mathcal{B}^n$ which is given by $\mathfrak{G}^n := (\mathsf{P}_{X^nY^n}, Q^{\wedge n})$ where

$$\mathsf{P}_{X^nY^n}(x^n, y^n) := \prod_{i=1}^{n} \mathsf{P}_{XY}(x_i, y_i), \quad \text{and}$$

$$Q^{\wedge n}(x^n, y^n, a^n, b^n) := \bigwedge_{i=1}^{n} Q(x_i, y_i, a_i, b_i).$$

If a strategy is given, the distribution $\mathsf{P}_{X^nY^nA^nB^n}$ of queries and answers is defined in the obvious way. We further define, for all i, the event W_i which occurs if the ith subgame is won.

Definition 2.4. For a game $\mathfrak{G}^n$ and a strategy (h_a, h_b) the distribution $\mathsf{P}_{X^nY^nA^nB^n}$ over $\mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{A}^n \times \mathcal{B}^n$ is given by

$$\mathsf{P}_{X^nY^nA^nB^n}(x^n, y^n, a^n, b^n) := \begin{cases} \mathsf{P}_{X^nY^n}(x^n, y^n) & \text{if } h_a(x^n) = a^n \text{ and } h_b(y^n) = b^n \\ 0 & \text{otherwise.} \end{cases}$$

Further, W^n is the tuple of events $(W_1, \ldots, W_n)$ where $W_i :\Longleftrightarrow Q(X_i, Y_i, A_i, B_i)$.

We prove the following version of the Parallel Repetition Theorem.

Theorem 2.5 (Parallel Repetition Theorem)**.** *For any game $\mathfrak{G}$ with value $v := v(\mathfrak{G})$ and any integer n:*

$$v(\mathfrak{G}^n) \leq \left(1 - \frac{(1-v)^3}{6000}\right)^{\frac{n}{\log(|\mathcal{A}||\mathcal{B}|)}}.$$

The constant 6000 could be improved by a more careful analysis (we will not optimize constants which would improve it during the proof). However, we do not know whether the 3 in the exponent can be reduced in the general case (as remarked above, it was reduced to 2 for some games by Rao [31]).

In [29] it is shown that in Raz's proof the term $\log(|\mathcal{A}||\mathcal{B}|)$ in the exponent can be reduced to the maximum of the logarithm of the partition number of $Q(x, y, \cdot, \cdot)$. As shown by Beame [4], the argument can be adapted to work with the proof given here. We give a slightly different argument in Section 8 which shows how the term can be reduced to a quantity which is a lower bound on the logarithm of the partition number.

3 Proof sketch

Fix an arbitrary game $\mathfrak{G}$, its n-fold parallel repetition $\mathfrak{G}^n$, and a strategy h_a, h_b for $\mathfrak{G}^n$. With the notation from Definition 2.4, the parallel repetition theorem is simply an upper bound on $\Pr[W_1 \wedge \cdots \wedge W_n]$. To get such an upper bound, we show that for arbitrary indices $i_1, \ldots, i_m$ there exists an index j such that

$$\Pr[W_j | W_{i_1} \wedge \cdots \wedge W_{i_m}] \leq v(\mathfrak{G}) + \varepsilon, \tag{3.1}$$

where ε depends on m, n, $\log(|\mathcal{A}||\mathcal{B}|)$, and $\Pr[W_{i_1} \wedge \cdots \wedge W_{i_m}]$ (this is Lemma 6.5). From (3.1) a simple induction gives the parallel repetition theorem, thus we now concentrate on the proof of (3.1).

Locally computable embeddings In order to prove (3.1) we define the distribution

$$\mathsf{P}_{\widetilde{X}^n\widetilde{Y}^n} := \mathsf{P}_{X^nY^n|W_{i_1}\wedge\cdots\wedge W_{i_m}} \tag{3.2}$$

(i. e., the distribution of the message which the referee sends to Alice and Bob conditioned on the event that the games i_1 to i_m are won).

We show (Lemma 6.4) that for some j the following can be achieved by Alice and Bob without communication and using shared randomness only:

1. Alice, on input x, produces a tuple $\bar{x}^n$ with $\bar{x}_j = x$.

2. Bob, on input y, produces a tuple $\bar{y}^n$ with $\bar{y}_j = y$.

3. Let $\mathsf{P}_{\overline{X}^n\overline{Y}^n}$ be the resulting joint distribution of the tuples $(\bar{x}^n, \bar{y}^n)$, assuming that (x,y) is chosen according to P_{XY}. Then
$$\|\mathsf{P}_{\overline{X}^n\overline{Y}^n} - \mathsf{P}_{\widetilde{X}^n\widetilde{Y}^n}\| \leq \varepsilon\,.$$

If 1-3 holds, we say that "(X,Y) can be $1-\varepsilon$-embedded into $(\widetilde{X}^n, \widetilde{Y}^n)$ with $(\widetilde{X}_j, \widetilde{Y}_j) = (X,Y)$ by local computation."

If such an embedding is given, we can consider the following strategy for the initial game $\mathfrak{G}$: Alice and Bob embed their inputs (X,Y) in $(\widetilde{X}^n, \widetilde{Y}^n)$ with $(\widetilde{X}_j, \widetilde{Y}_j) = (X,Y)$, and answer with coordinate j of $h_a(\widetilde{X}^n)$ and $h_b(\widetilde{Y}^n)$. This strategy wins with probability at least $\Pr[W_j|W_{i_1}\wedge\cdots\wedge W_{i_m}] - \varepsilon$. Since no strategy for the initial game has higher winning probability than $v(\mathfrak{G})$ this implies (3.1).

We remark that a necessary condition for such an embedding to exist is that

$$\|\mathsf{P}_{XY} - \mathsf{P}_{\widetilde{X}_j\widetilde{Y}_j}\| \leq \varepsilon\,, \tag{3.3}$$

and indeed this follows from Lemma 4.1 for $U_j = (X_j, Y_j)$ (of course this condition is not a sufficient one).

Constructing an embedding We now give a more detailed explanation how Alice and Bob can embed (X,Y) into $(\widetilde{X}^n, \widetilde{Y}^n)$ with $(\widetilde{X}_j, \widetilde{Y}_j) = (X,Y)$. For this, given values (x,y) distributed according to P_{XY}, Alice and Bob proceed as follows:

1. Alice and Bob use shared randomness to produce queries and responses for all the games won, i. e., values $(x_{i_1}, y_{i_1}, a_{i_1}, b_{i_1})$ to $(x_{i_m}, y_{i_m}, a_{i_m}, b_{i_m})$. Here, Alice and Bob *both* produce *all* these values consistently (i. e., they ensure that they produce the *same* actual values).

2. For every index $i \notin \{i_1, \ldots, i_m, j\}$, Alice and Bob examine a shared random bit d_i. If $d_i = 1$ both locally produce x_i, otherwise both locally produce y_i. Again, Alice and Bob both produce all these values consistently.

3. Using individual randomness, Alice and Bob locally expand their information such that Alice gets x^n and Bob y^n.

In steps 1 and 2 we have to take care of two things: first, the values produced should be distributed according to the respective marginals of the distribution $\mathsf{P}_{\widetilde{A}^n\widetilde{B}^n\widetilde{X}^n\widetilde{Y}^n|\widetilde{X}_j=x\wedge\widetilde{Y}_j=y}$ (where $\mathsf{P}_{\widetilde{A}^n\widetilde{B}^n\widetilde{X}^n\widetilde{Y}^n}$ is defined analogously to (3.2)). Second, Alice and Bob should produce *equal values* (otherwise the resulting random variables $(\overline{X}^n,\overline{Y}^n)$ will not have the correct overall distribution).

For step 1 achieving both is simple: it follows from Corollary 4.3 that Alice and Bob can choose the values $(x_{i_1},y_{i_1},a_{i_1},b_{i_1}),\dots,(x_{i_m},y_{i_m},a_{i_m},b_{i_m})$ independently of (x,y) according to $\mathsf{P}_{\widetilde{X}_{i_1}\widetilde{Y}_{i_1}\widetilde{A}_{i_1}\widetilde{B}_{i_1}\cdots\widetilde{X}_{i_m}\widetilde{Y}_{i_m}\widetilde{A}_{i_m}\widetilde{B}_{i_m}}$. Using shared randomness this can be done such that both get the same tuple.

The second step is harder, as in this case the values cannot be chosen independently of (x_j,y_j) anymore.[2] However, let $\widetilde{S}$ be the random variables which Alice and Bob produce in Step 2. It will follow from Corollary 4.3 that

$$\left\|\mathsf{P}_{XY}\mathsf{P}_{\widetilde{S}|\widetilde{X}_j}-\mathsf{P}_{XY\widetilde{S}}\right\| \quad\text{and}\quad \left\|\mathsf{P}_{XY}\mathsf{P}_{\widetilde{S}|\widetilde{Y}_j}-\mathsf{P}_{XY\widetilde{S}}\right\|$$

are both small, and Lemma 5.2 implies that this is sufficient to generate $\widetilde{S}$ locally.

In fact, Corollary 4.3 and Lemma 5.2 are strong enough to do steps 1 and 2 at the same time, and thus these steps are done simultaneously in the proof of Lemma 6.4.

Step 3 will be simpler to implement, since conditioned on the values already generated, the values Alice generates are independent of the values Bob generates. To see this, consider the following experiment. Pick (X^n,Y^n) according to $(\mathsf{P}_{XY})^n$, then, for each coordinate reveal either X or Y. Clearly, given this information, the non-revealed X_i are independent of the non-revealed Y_i. Furthermore, this does not change even if additionally coordinates $i_1,\dots,i_m$ of $h_a(X^n)$ and $h_b(Y^n)$ are revealed to both, which is the situation before step 3 (actually, one needs to take a little more care because Alice does not know exactly the same values as Bob: Alice knows X_j while Bob knows Y_j—see Lemma 5.4 for the exact conditions).

4 Conditioned distributions

The following lemma is essentially Claim 5.1 in Raz's paper [32] (and we use the proof given there). It states that if random variables U_i are chosen independently, then conditioning on an event does not change the individual distributions a lot on average, unless the event has very low probability.

Lemma 4.1. *Let $\mathsf{P}_{U^k}:=\mathsf{P}_{U_1}\cdots\mathsf{P}_{U_k}$ be a probability distribution over $\mathcal{U}^k$, W an event. Then,*

$$\Pr[W]\le 2^{-\sum_{j=1}^{k}(\|\mathsf{P}_{U_j|W}-\mathsf{P}_{U_j}\|)^2}. \tag{4.1}$$

As an example, let U_i be uniform and independent bits and W be the event that at least $k(\frac{1}{2}+\varepsilon)$ of these bits are one. Then $\|P_{U_i|W}-\mathsf{P}_{U_i}\|\ge\varepsilon$ and the lemma states that $\Pr[W]\le 2^{-k\varepsilon^2}$, which is a version of Chernoff's inequality (note that this implies that Lemma 4.1 is almost tight; see, for example, [22]).

[2] The values d_i can be chosen independently, but *not* the values of the x_i or the y_i. We quickly explain why this is impossible in general. Assume that the random variables X and Y contain a shared bit B. The game $\mathfrak{G}^n$ and the strategy (h_a,h_b) may be such that Alice and Bob win subgame i_1 in case $B_1\oplus\cdots\oplus B_n=0$. Generating the values independently of (x,y) would now produce a distribution with statistical distance at least $\frac{1}{2}$ from the target distribution. Therefore, a bit which is contained in both x and y *must* be considered when generating the values of x_i and y_i.

Using $(\sum_{j=1}^{k} a_j)^2 \leq k \sum_{j=1}^{k} a_j^2$ one checks that (4.1) implies

$$\sum_{j=1}^{k} \|\mathsf{P}_{U_j|W} - \mathsf{P}_{U_j}\| \leq \sqrt{k \log\left(\frac{1}{\Pr[W]}\right)}\,, \tag{4.2}$$

which is the form we use later.

The proof of Lemma 4.1 uses the *relative entropy* $D(\mathsf{P}_S \| \mathsf{P}_T)$, which, for two distributions P_S and P_T over the same set $\mathcal{S}$, is defined as

$$D(\mathsf{P}_S \| \mathsf{P}_T) := \sum_{s \in \mathcal{S}} \mathsf{P}_S(s) \log\left(\frac{\mathsf{P}_S(s)}{\mathsf{P}_T(s)}\right) \tag{4.3}$$

(which is ∞ by convention if there is $s \in \mathcal{S}$ with $\mathsf{P}_S(s) > 0 = \mathsf{P}_T(s)$). This quantity satisfies $D(\mathsf{P}_S \| \mathsf{P}_T) \geq \left(\|\mathsf{P}_S - \mathsf{P}_T\|\right)^2$ (see [12, Lemma 11.6.1]). Furthermore, we have the following inequality, which is well known, but for which we could not find a written proof in the literature.

Lemma 4.2. *For any distributions $\mathsf{P}_{U^k} = \mathsf{P}_{U_1} \cdots \mathsf{P}_{U_k}$ and P_{V^k} over a set $\mathcal{U}^k$ we have*

$$\sum_{j=1}^{k} D(\mathsf{P}_{V_j} \| \mathsf{P}_{U_j}) \leq D(\mathsf{P}_{V^k} \| \mathsf{P}_{U^k})\,. \tag{4.4}$$

Proof. We prove the case $k = 2$; the general case follows by induction. We get

$$\begin{aligned} D(\mathsf{P}_{V_1V_2} \| \mathsf{P}_{U_1}\mathsf{P}_{U_2}) &= \sum_{u_1,u_2} \mathsf{P}_{V_1V_2}(u_1,u_2) \log\left(\frac{\mathsf{P}_{V_1}(u_1)}{\mathsf{P}_{U_1}(u_1)}\right) + \sum_{u_1,u_2} \mathsf{P}_{V_1V_2}(u_1,u_2) \log\left(\frac{\mathsf{P}_{V_2}(u_2)}{\mathsf{P}_{U_2}(u_2)}\right) \\ &\quad + \sum_{u_1,u_2} \mathsf{P}_{V_1V_2}(u_1,u_2) \log\left(\frac{\mathsf{P}_{V_1V_2(u_1,u_2)}}{\mathsf{P}_{V_1}(u_1)\mathsf{P}_{V_2}(u_2)}\right) \\ &= D(\mathsf{P}_{V_1} \| \mathsf{P}_{U_1}) + D(\mathsf{P}_{V_2} \| \mathsf{P}_{U_2}) + D(\mathsf{P}_{V_1V_2} \| \mathsf{P}_{V_1}\mathsf{P}_{V_2})\,, \end{aligned}$$

and $D(\mathsf{P}_{V_1V_2} \| \mathsf{P}_{V_1}\mathsf{P}_{V_2}) \geq 0$ (see [12, Theorem 2.6.3]). □

We can now give the proof of Lemma 4.1.

Proof of Lemma 4.1. Using the above we get

$$\begin{aligned} \sum_{j=1}^{k} \left(\|\mathsf{P}_{U_j|W} - \mathsf{P}_{U_j}\|\right)^2 &\leq \sum_{j=1}^{k} D(\mathsf{P}_{U_j|W} \| \mathsf{P}_{U_j}) \\ &\leq D(\mathsf{P}_{U^k|W} \| \mathsf{P}_{U^k}) \end{aligned}$$

$$
\begin{aligned}
&= \sum_{u^k} \mathsf{P}_{U^k|W}(u^k) \log\Big(\frac{\mathsf{P}_{U^k|W}(u^k)}{\mathsf{P}_{U^k}(u^k)}\Big) \\
&= \sum_{u^k} \mathsf{P}_{U^k|W}(u^k) \log\Big(\frac{\Pr[W|U^k=u^k]}{\Pr[W]}\Big) \\
&= \log\Big(\frac{1}{\Pr[W]}\Big) + \sum_{u^k} \mathsf{P}_{U^k|W}(u^k) \log\big(\Pr[W|U^k=u^k]\big) \\
&\le \log\Big(\frac{1}{\Pr[W]}\Big). \qquad \square
\end{aligned}
$$

We next give a slight extension of Lemma 4.1 (this makes it simpler to apply later). First, the U_j are independent given the value of an additional random variable T. Second, an arbitrary third random variable V with bounded alphabet size gives side information about U_j. Then, choosing U_j without considering the fact that an event W happened and ignoring V does not change the distribution of U_j too much on average. For the notation in the following corollary we refer to Section 2.1, equation (2.1) and the subsequent remarks.

Corollary 4.3. *Let $\mathsf{P}_{TU^kV} := \mathsf{P}_T \mathsf{P}_{U_1|T} \mathsf{P}_{U_2|T} \cdots \mathsf{P}_{U_k|T} \mathsf{P}_{V|TU^k}$ be a probability distribution over $\mathcal{T} \times \mathcal{U}^k \times \mathcal{V}$, W be an event. Then,*

$$
\sum_{j=1}^{k} \Big\| \mathsf{P}_{TU_jV|W} - \mathsf{P}_{TV|W} \mathsf{P}_{U_j|T} \Big\| \le \sqrt{k} \sqrt{\log(|\mathcal{V}^*|) + \log\Big(\frac{1}{\Pr[W]}\Big)},
$$

where $\mathcal{V}^* := \{v \in \mathcal{V} | \mathsf{P}_{V|W}(v) > 0\}$.

The proof is essentially an application of Jensen's inequality on Lemma 4.1.

Proof. Fix a pair $(t,v) \in \mathcal{T} \times \mathcal{V}$ and consider the distributions $\mathsf{P}_{U^k|T=t,V=v,W}$ and $\mathsf{P}_{U^k|T=t}$. We apply Lemma 4.1 (in the form given by (4.2)) on these distributions (with the event $(V{=}v) \wedge W$) and get

$$
\begin{aligned}
\sum_{j=1}^{k} \Big\| \mathsf{P}_{TU_jV|W} - \mathsf{P}_{TV|W} \mathsf{P}_{U_j|T} \Big\| &= \sum_{(t,v):\mathsf{P}_{TV|W}(t,v)>0} \mathsf{P}_{TV|W}(t,v) \cdot \sum_{j=1}^{k} \Big\| \mathsf{P}_{U_j|T=t,V=v,W} - \mathsf{P}_{U_j|T=t} \Big\| \\
&\le \sum_{(t,v):\mathsf{P}_{TV|W}(t,v)>0} \mathsf{P}_{TV|W}(t,v) \sqrt{k \log\Big(\frac{1}{\Pr[W \wedge V=v | T=t]}\Big)} \\
&\le \sqrt{k \log\Big(\sum_{(t,v):\mathsf{P}_{TV|W}(t,v)>0} \mathsf{P}_{TV|W}(t,v) \frac{1}{\Pr[W \wedge V=v|T=t]}\Big)}, \qquad (4.5)
\end{aligned}
$$

where the last inequality is Jensen's inequality applied to the function $\sqrt{\log(\cdot)}$, concave on $[1,\infty)$. We compute

$$\begin{aligned}\sum_{(t,v):\mathsf{P}_{TV|W}(t,v)>0} \mathsf{P}_{TV|W}(t,v)\frac{1}{\Pr[W\wedge V=v|T=t]} &= \sum_{(t,v):\mathsf{P}_{TV|W}(t,v)>0} \frac{\Pr[T=t\wedge V=v|W]}{\Pr[W\wedge V=v|T=t]}\\ &= \sum_{(t,v):\mathsf{P}_{TV|W}(t,v)>0} \frac{\Pr[T=t\wedge V=v\wedge W]\Pr[T=t]}{\Pr[W]\Pr[V=v\wedge T=t\wedge W]}\\ &= \sum_{(t,v):\mathsf{P}_{TV|W}(t,v)>0} \frac{\Pr[T=t]}{\Pr[W]} = \frac{|\mathcal{V}^*|}{\Pr[W]}\,.\end{aligned}$$

Inserting this into (4.5) completes the proof. □

5 Embedding by local computation

We next study under what conditions random variables can be embedded into other random variables by local computations. The intent is that Alice holds a random variable X, Bob holds Y, and they would like to produce S and T, respectively, using the shared randomness R.

Definition 5.1 (Embeddable)**.** For two distributions $\mathsf{P}_{X_0Y_0}$ and $\mathsf{P}_{X_1SY_1T}$ we say that (X_0,Y_0) is $(1-\varepsilon)$-embeddable in (X_1S,Y_1T) with $(X_1,Y_1)=(X_0,Y_0)$ if there exists a probability measure P_R over a set $\mathcal{R}$ and functions $f_A:\mathcal{X}\times\mathcal{R}\to\mathcal{S}$, $f_B:\mathcal{Y}\times\mathcal{R}\to\mathcal{T}$, such that

$$\left\|\mathsf{P}_{X_0Y_0}\mathsf{P}_{F_AF_B|XY}-\mathsf{P}_{X_1Y_1ST}\right\|\le\varepsilon\,,$$

where $\mathsf{P}_{F_AF_B|X=xY=y}$ is the distribution defined by the random variable $(f_A(x,R),f_B(y,R))$.

The following lemma gives a condition under which (X,Y) is embeddable in (XS,YS).

Lemma 5.2. *Let a distribution P_{SXY} be given. If*

$$\|\mathsf{P}_{SXY}-\mathsf{P}_{XY}\mathsf{P}_{S|X}\|\le\varepsilon_1 \tag{5.1}$$

and

$$\|\mathsf{P}_{SXY}-\mathsf{P}_{XY}\mathsf{P}_{S|Y}\|\le\varepsilon_2\,, \tag{5.2}$$

then (X,Y) is $(1-2\varepsilon_1-2\varepsilon_2)$-embeddable[3] *in (XS,YS).*

Even if $\varepsilon_1=\varepsilon_2=0$, equations (5.1) and (5.2) do not imply that S is independent of X and Y. For example, if X and Y contain the same uniform random bit, then S can depend on this bit. However, if $\varepsilon_1=\varepsilon_2=0$ the lemma is obviously true: Alice uses shared randomness to choose S according to $\mathsf{P}_{S|X=x}$ (more concretely: Alice chooses a uniform random real $\rho\in[0,1]$ and uses the smallest element s for

[3]It is understood that the embedding satisfies $(X,Y)=(X,Y)$, i. e., that the original random variables will result if from the resulting (XS,YS) the S-part is omitted.

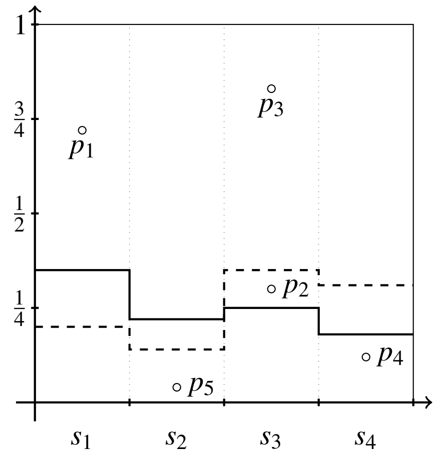

Figure 1: Proof sketch of Lemma 5.2: Alice and Bob use the shared randomness to pick an infinite sequence of points $\{p_i\}_{i\geq 0}$ in the plane. Both consider the first point which falls below their density function, and pick the corresponding element. Here, Alice (dashed line) picks s_3, since p_2 is the first point below the dashed line. Bob (solid line) picks s_4, due to p_4. The probability that Alice and Bob pick different values s_i is approximately proportional to the area between the lines, which is twice the statistical distance.

which the cumulative distribution function $\sum_{s'\leq s} \mathsf{P}_{S|X=x}(s')$ is larger than ρ). Since Bob has the same distribution $\mathsf{P}_{S|Y=y}$ he will find the same value if he uses the same shared randomness.

In case $\varepsilon_1 > 0$ and $\varepsilon_2 > 0$, we have to overcome the following problem: $\mathsf{P}_{S|Y=y}$ is unknown to Alice (since y is unknown to Alice), and analogously $\mathsf{P}_{S|X=x}$ is unknown to Bob. The solution is depicted graphically in Figure 1 for an alphabet of size 4: Alice and Bob paint their density function on a piece of paper (where every element gets assigned the same width), then use the shared randomness to sample points uniformly in the rectangle until one of the points is below their density. They pick the element whose label is below the point.

Note that they will get a different output only if the first element which is below any of the two lines is above the other line. This happens with probability

$$\frac{\|\mathsf{P}_{S|X=x}-\mathsf{P}_{S|Y=y}\|}{1+\|\mathsf{P}_{S|X=x}-\mathsf{P}_{S|Y=y}\|}\,.$$

We formalize this in the following proof.

Proof of Lemma 5.2. Let $\mathcal{R} := (\mathcal{S}\times[0,1])^\infty$ be the set of infinite sequences over $\mathcal{S}\times[0,1]$. For a fixed x,y and a sequence $r := \{(s_i,\rho_i)\}_{i\geq 0} \in \mathcal{R}$, we define $f_A(x,r) := s_i$ if i is the smallest index for which $\mathsf{P}_{S|X=x}(s_i) > \rho_i$. Analogously, we define $f_B(y,r) := s_j$ if j is the smallest index with $\mathsf{P}_{S|Y=y}(s_j) > \rho_j$ and[4] $f_{AB}(x,y,r) := s_k$ if k is the smallest index with $\mathsf{P}_{S|X=xY=y}(s_k) > \rho_k$. If no such index exists, the corresponding function is defined in an arbitrary way (this happens with probability 0).

[4]The use of f_{AB} in order to simplify the analysis was suggested by Anup Rao.

Let $\mathsf{P}_{XYF_AF_BF_{AB}}$ be the joint distribution of $(x,y,f_A(x,r),f_B(y,r),f_{AB}(x,y,r))$ where (x,y) is chosen according to P_{XY} and r uniformly from $\mathcal{R}$. We have $\mathsf{P}_{F_{AB}|X=xY=y} = \mathsf{P}_{S|X=xY=y}$, $\mathsf{P}_{F_A|X=x} = \mathsf{P}_{S|X=x}$ and $\mathsf{P}_{F_B|Y=y} = \mathsf{P}_{S|Y=y}$, since these equalities hold conditioned on the event that the corresponding function accepts in round i, for any fixed i.

Further, we have $\Pr[F_A = F_{AB}|X=x,Y=y] \geq 1-2\|\mathsf{P}_{F_A|X=x} - \mathsf{P}_{F_{AB}|X=xY=y}\|$: the two values F_A, F_{AB} are equal if $\rho_j < \min(\mathsf{P}_{F_A|X=x}(s_j), \mathsf{P}_{F_{AB}|X=xY=y}(s_j))$ for the smallest j for which $\rho_j < \max(\mathsf{P}_{F_A|X=x}(s_j), \mathsf{P}_{F_{AB}|X=xY=y}(s_j))$ is satisfied. This happens with probability

$$\begin{aligned}\frac{\sum_s \min(\mathsf{P}_{F_A|X=x}(s_j), \mathsf{P}_{F_{AB}|X=xY=y}(s_j))}{\sum_s \max(\mathsf{P}_{F_A|X=x}(s_j), \mathsf{P}_{F_{AB}|X=xY=y}(s_j))} &= \frac{1-\|\mathsf{P}_{F_A|X=x} - \mathsf{P}_{F_{AB}|X=xY=y}\|}{1+\|\mathsf{P}_{F_A|X=x} - \mathsf{P}_{F_{AB}|X=xY=y}\|} \\ &\geq 1-2\|\mathsf{P}_{F_A|X=x} - \mathsf{P}_{F_{AB}|X=xY=y}\|\,.\end{aligned}$$

This yields $\Pr[F_A = F_{AB}] \geq 1-2\varepsilon_1$, and analogously we get $\Pr[F_B = F_{AB}] \geq 1-2\varepsilon_2$, and thus $\Pr[F_A = F_B = F_{AB}] \geq 1-2\varepsilon_1-2\varepsilon_2$. This implies

$$\|\mathsf{P}_{XYSS} - \mathsf{P}_{XY}\mathsf{P}_{F_AF_B|XY}\| = \|\mathsf{P}_{XYF_{AB}F_{AB}} - \mathsf{P}_{XYF_AF_B}\| \geq 1-2\varepsilon_1-2\varepsilon_2\,. \qquad \square$$

In the following corollary, the input distribution is changed slightly. This makes it a bit easier to apply later.

Corollary 5.3. *Let distributions* P_{SXY} *and* $\mathsf{P}_{X_0Y_0}$ *be given. If*

$$\|\mathsf{P}_{SXY} - \mathsf{P}_{X_0Y_0}\mathsf{P}_{S|X}\| \leq \varepsilon_1 \tag{5.3}$$

and

$$\|\mathsf{P}_{SXY} - \mathsf{P}_{X_0Y_0}\mathsf{P}_{S|Y}\| \leq \varepsilon_2\,, \tag{5.4}$$

then (X_0,Y_0) *is* $(1-3\varepsilon_1-2\varepsilon_2)$*-embeddable*[5] *in* (XS,YS) *with* $(X,Y)=(X_0,Y_0)$.

Proof. From (5.3) we get $\|\mathsf{P}_{XY} - \mathsf{P}_{X_0Y_0}\| \leq \varepsilon_1$. One can now find a joint distribution $\mathsf{P}_{XYX_0Y_0}$ with $\Pr[(X,Y)=(X_0,Y_0)] \geq 1-\varepsilon_1$. The corollary now follows by applying f_A and f_B from Lemma 5.2. $\square$

Random variables S,T,U form a Markov chain, written $S \leftrightarrow T \leftrightarrow U$, if $\mathsf{P}_{STU} = \mathsf{P}_T\mathsf{P}_{S|T}\mathsf{P}_{U|T}$ (i. e., if given T the random variable U is independent of S). The following lemma is essentially Lemma 4.1 in Raz's paper.

Lemma 5.4. *Let* P_{XYST} *be any distribution. If*

$$S \leftrightarrow X \leftrightarrow YT \qquad \text{and} \qquad XS \leftrightarrow Y \leftrightarrow T$$

then (X,Y) *is* 1*-embeddable in* (XS,YT).

Proof. Using individual (non-shared) randomness, Alice computes S according to $\mathsf{P}_{S|X=x}$ and Bob computes T according to $\mathsf{P}_{T|Y=y}$. Since

$$\mathsf{P}_{STXY} = \mathsf{P}_{XY}\mathsf{P}_{S|XY}\mathsf{P}_{T|SXY} = \mathsf{P}_{XY}\mathsf{P}_{S|X}\mathsf{P}_{T|Y} \tag{5.5}$$

this gives the correct (global) distribution. $\square$

[5] The statement could be made symmetric: (X_0,Y_0) is $(1-2\varepsilon_1-2\varepsilon_2-\min(\varepsilon_1,\varepsilon_2))$-embeddable.

6 Embeddings for games

Given a game $\mathfrak{G}$ and its n-fold parallel repetition, we now show that (X,Y) can be embedded into $(\widetilde{X}^n,\widetilde{Y}^n)$, where $\mathsf{P}_{\widetilde{X}^n\widetilde{Y}^n} := \mathsf{P}_{X^nY^n|W_{k+1}\wedge\cdots\wedge W_n}$.

We need the following simple fact on statistical distance.

Claim 6.1. *Let P_{Z_0} and P_{Z_1} be distributions over $\mathcal{Z}$. Let $\mathcal{S} \subseteq \mathcal{Z}$ be such that* $\Pr[Z_0 \in \mathcal{S}] = \Pr[Z_1 \in \mathcal{S}] = \frac{1}{2}$. *Then,*

$$\|\mathsf{P}_{Z_0|Z_0\in\mathcal{S}} - \mathsf{P}_{Z_1|Z_1\in\mathcal{S}}\| \le 2\|\mathsf{P}_{Z_0} - \mathsf{P}_{Z_1}\|\,.$$

Proof. Use Lemma 2.1, (2.3). □

Also, we need the following statements about Markov chains.

Claim 6.2. *Let $\mathsf{P}_{X_0Y_0}\mathsf{P}_{X_1Y_1}$ be a distribution over $\mathcal{X}_0\times\mathcal{Y}_0\times\mathcal{X}_1\times\mathcal{Y}_1$, $f:\mathcal{X}_0\times\mathcal{X}_1\to\mathcal{U}$ and $g:\mathcal{Y}_0\times\mathcal{Y}_1\to\mathcal{V}$ be arbitrary. Then,*

$$X_0X_1 \leftrightarrow X_0f(X_0,X_1)Y_1g(Y_0,Y_1) \leftrightarrow Y_0Y_1\,. \tag{6.1}$$

Proof. It is sufficient to show this for all possible values $x_0 \in \mathcal{X}_0$ and $y_1 \in \mathcal{Y}_1$. Let

$$\mathsf{P}_{\widetilde{Y}_0\widetilde{X}_1} := \mathsf{P}_{Y_0X_1|X_0=x_0Y_1=y_1} = \mathsf{P}_{Y_0|X_0=x_0}\mathsf{P}_{X_1|Y_1=y_1}\,.$$

In this case, (6.1) reduces to

$$\widetilde{X}_1 \leftrightarrow f(x_0,\widetilde{X}_1)g(\widetilde{Y}_0,y_1) \leftrightarrow \widetilde{Y}_0\,.$$

Since $\widetilde{X}_1$ and $\widetilde{Y}_0$ are independent this is obvious. □

Claim 6.3. *Let P_{TUV} be a distribution over $\mathcal{T}\times\mathcal{U}\times\mathcal{V}$ and W an event with*

$$T \leftrightarrow U \leftrightarrow V,$$
$$W \leftrightarrow U \leftrightarrow TV\,.$$

Then, for $\mathsf{P}_{\widetilde{T}\widetilde{U}\widetilde{V}} := \mathsf{P}_{TUV|W}$ we have $\widetilde{T} \leftrightarrow \widetilde{U} \leftrightarrow \widetilde{V}$.

Proof.

$$\begin{aligned}
\mathsf{P}_{\widetilde{T}\widetilde{U}\widetilde{V}}(t,u,v) &= \mathsf{P}_{TUV|W}(t,u,v)\\
&= \mathsf{P}_{U|W}(u)\mathsf{P}_{TV|U=u,W}(t,v)\\
&= \mathsf{P}_{U|W}(u)\mathsf{P}_{TV|U=u}(t,v)\\
&= \mathsf{P}_{U|W}(u)\mathsf{P}_{T|U=u}(t)\mathsf{P}_{V|U=u}(v)\\
&= \mathsf{P}_{U|W}(u)\mathsf{P}_{T|U=u,W}(t)\mathsf{P}_{V|U=u,W}(v)\,.
\end{aligned}$$

□

Lemma 6.4. *Let a game $\mathfrak{G}^n = (Q^n, (\mathsf{P}_{XY})^n)$, a strategy (h_a, h_b), and $k \le n$ be given. Let*

$$\mathsf{P}_{\widetilde{X}^n\widetilde{Y}^n} := \mathsf{P}_{X^nY^n|W_{k+1}\wedge\cdots\wedge W_n}.$$

Then, for $1 \le j \le k$, there exists $\varepsilon_j \ge 0$ such that (X,Y) is $(1-\varepsilon_j)$-embeddable in $(\widetilde{X}^n, \widetilde{Y}^n)$ with $(\widetilde{X}_j, \widetilde{Y}_j) = (X,Y)$ and

$$\sum_{j=1}^{k} \varepsilon_j \le 15\sqrt{k}\sqrt{(n-k)\log(|\mathcal{A}|\,|\mathcal{B}|) + \log\Big(\frac{1}{\Pr[W_{k+1}\wedge\cdots\wedge W_n]}\Big)}. \tag{6.2}$$

Proof. As described in Definition 2.4 we consider the distribution $\mathsf{P}_{X^nY^nA^nB^n}$, the corresponding random variables, and the events W^n. Additionally, we let $D_1, \dots, D_k$ be uniform and independent bits. For $1 \le j \le k$ we define

$$U_j := \begin{cases} X_j & \text{if } D_j = 0 \\ Y_j & \text{otherwise} \end{cases}$$

and

$$\overline{U}_j := \begin{cases} Y_j & \text{if } D_j = 0 \\ X_j & \text{otherwise.} \end{cases}$$

Also, we set

$$T := (X_{k+1}, \dots, X_n, Y_{k+1}, \dots, Y_n, D^k, \overline{U}^k)\,, \tag{6.3}$$
$$V := (A_{k+1}, \dots, A_n, B_{k+1}, \dots, B_n)\,, \tag{6.4}$$

and define the event $W := W_{k+1} \wedge \cdots \wedge W_n$.

From Corollary 4.3 we get

$$\sum_{j=1}^{k} \Big\| \mathsf{P}_{TU_jV|W} - \mathsf{P}_{TV|W}\mathsf{P}_{U_j|T} \Big\| \le \varepsilon_{\mathrm{Tot}}\,, \tag{6.5}$$

where we set

$$\varepsilon_{\mathrm{Tot}} := \sqrt{k}\sqrt{(n-k)\log(|\mathcal{A}||\mathcal{B}|) + \log\Big(\frac{1}{\Pr[W]}\Big)} \tag{6.6}$$

(we applied Corollary 4.3 using $|\mathcal{V}^*| \le |\mathcal{V}|$).

In (6.5), we condition on both sides on the event $D_j = 0$, which is, on both sides, a restriction on a subset which has probability $\frac{1}{2}$. Claim 6.1 implies

$$\sum_{j=1}^{k} \Big\| \mathsf{P}_{TU_jV|W\wedge(D_j=0)} - \mathsf{P}_{TV|W\wedge(D_j=0)}\mathsf{P}_{U_j|T} \Big\| \le 2\varepsilon_{\mathrm{Tot}}\,, \tag{6.7}$$

where we do not need to condition on $D_j = 0$ in $\mathsf{P}_{U_j|T}$ since this is included in the given t anyhow; in fact we can now write $\mathsf{P}_{X_j|Y_j}$ instead of $\mathsf{P}_{U_j|T}$.

For a fixed j, define the random variable

$$T^{(\setminus j)} := (X_{k+1}, \dots, X_n, Y_{k+1}, \dots, Y_n, \\ D_1, \dots, D_{j-1}, D_{j+1}, \dots, D_k, \\ \overline{U}_1, \dots, \overline{U}_{j-1}, \overline{U}_{j+1}, \dots, \overline{U}_k)\,. \tag{6.8}$$

With this notation (6.7) is equivalent to

$$\sum_{j=1}^{k} \left\| \mathsf{P}_{T^{(\setminus j)} X_j Y_j V | W \wedge (D_j=0)} - \mathsf{P}_{T^{(\setminus j)} Y_j V | W \wedge (D_j=0)} \mathsf{P}_{X_j|Y_j} \right\| \leq 2\varepsilon_{\text{Tot}}\,. \tag{6.9}$$

But now nothing depends on $D_j = 0$ anymore, so this also means

$$\sum_{j=1}^{k} \left\| \mathsf{P}_{T^{(\setminus j)} X_j Y_j V | W} - \mathsf{P}_{T^{(\setminus j)} Y_j V | W} \mathsf{P}_{X_j|Y_j} \right\| \leq 2\varepsilon_{\text{Tot}}\,. \tag{6.10}$$

We set $S := (T^{(\setminus j)}, V)$ and define the probability distribution

$$\mathsf{P}_{\widetilde{S}\widetilde{X}^n\widetilde{Y}^n} := \mathsf{P}_{SX^nY^n|W}\,. \tag{6.11}$$

With this, (6.10) becomes

$$\sum_{j=1}^{k} \left\| \mathsf{P}_{\widetilde{S}\widetilde{X}_j\widetilde{Y}_j} - \mathsf{P}_{\widetilde{S}\widetilde{Y}_j} \mathsf{P}_{X_j|Y_j} \right\| \leq 2\varepsilon_{\text{Tot}}\,, \tag{6.12}$$

or, equivalently

$$\sum_{j=1}^{k} \left\| \mathsf{P}_{\widetilde{S}\widetilde{X}_j\widetilde{Y}_j} - \mathsf{P}_{\widetilde{Y}_j} \mathsf{P}_{\widetilde{S}|\widetilde{Y}_j} \mathsf{P}_{X|Y} \right\| \leq 2\varepsilon_{\text{Tot}}\,. \tag{6.13}$$

Lemma 4.1 implies

$$\sum_{j=1}^{k} \left\| \mathsf{P}_{\widetilde{Y}_j} - \mathsf{P}_{Y} \right\| \leq \varepsilon_{\text{Tot}}\,, \tag{6.14}$$

and thus

$$\sum_{j=1}^{k} \left\| \mathsf{P}_{\widetilde{S}\widetilde{X}_j\widetilde{Y}_j} - \mathsf{P}_{XY} \mathsf{P}_{\widetilde{S}|\widetilde{Y}_j} \right\| \leq 3\varepsilon_{\text{Tot}}\,. \tag{6.15}$$

Symmetric reasoning yields

$$\sum_{j=1}^{k} \left\| \mathsf{P}_{\widetilde{S}\widetilde{X}_j\widetilde{Y}_j} - \mathsf{P}_{XY} \mathsf{P}_{\widetilde{S}|\widetilde{X}_j} \right\| \leq 3\varepsilon_{\text{Tot}}\,. \tag{6.16}$$

From (6.15) and (6.16), Corollary 5.3 implies that (X,Y) is $(1-\varepsilon_j)$-embeddable in $(\widetilde{X}_j\widetilde{S},\widetilde{Y}_j\widetilde{S})$ with $(\widetilde{X}_j,\widetilde{Y}_j)=(X,Y)$ and such that $\sum_{j=1}^{k}\varepsilon_j \le 15\varepsilon_{\mathrm{Tot}}$.

We next show that

$$X^k \leftrightarrow TV \leftrightarrow Y^k\,. \tag{6.17}$$

If the bits D^k and the values $X_{k+1},\dots,X_n, Y_{k+1},\dots,Y_n$ are fixed, this follows immediately from Claim 6.2. Since it holds for all these values, it must also hold overall.

From (6.17) we easily get

$$\begin{aligned} X^n &\leftrightarrow X_jS \leftrightarrow Y^nY_jS\\ X^nX_jS &\leftrightarrow Y_jS \leftrightarrow Y^n\,. \end{aligned}$$

Claim 6.3 yields

$$\widetilde{X}^n \leftrightarrow \widetilde{X}_j\widetilde{S} \leftrightarrow \widetilde{Y}^n\widetilde{Y}_j\widetilde{S} \tag{6.18}$$

$$\widetilde{X}^n\widetilde{X}_j\widetilde{S} \leftrightarrow \widetilde{Y}_j\widetilde{S} \leftrightarrow \widetilde{Y}^n\,. \tag{6.19}$$

Above we have seen that (X,Y) is embeddable in $(\widetilde{X}_j\widetilde{S},\widetilde{Y}_j\widetilde{S})$ with $(\widetilde{X}_j,\widetilde{Y}_j)=(X,Y)$. Lemma 5.4 together with (6.18) and (6.19) now implies that we can 1-locally embed this in $(\widetilde{X}^n\widetilde{X}_j\widetilde{S},\widetilde{Y}^n\widetilde{Y}_j\widetilde{S})$. Since Alice and Bob can then ignore part of the information constructed, this completes the proof. □

Lemma 6.5. *Let a game $\mathfrak{G}=(Q,\mathsf{P}_{XY})$, its n-fold repetition $\mathfrak{G}^n$, and a strategy (h_a,h_b) for $\mathfrak{G}^n$ be given. Let indices $i_1,\dots,i_m$ be given. Then there exists an index i_{m+1} such that*

$$\begin{aligned} &\Pr[W_{i_{m+1}}|W_{i_1}\wedge\cdots\wedge W_{i_m}]\\ &\qquad \le v(\mathfrak{G})+15\sqrt{\frac{1}{n-m}}\sqrt{m\log(|\mathcal{A}||\mathcal{B}|)+\log\Big(\frac{1}{\Pr[W_{i_1}\wedge\cdots\wedge W_{i_m}]}\Big)}\,. \end{aligned} \tag{6.20}$$

Proof. First, we can assume that the given indices i_ℓ, $1\le\ell\le m$, are pairwise different (otherwise we get a stronger statement). Given this we can even assume that $i_\ell=n-\ell+1$ by appropriately redefining the functions (h_a,h_b).

Define the distribution $\mathsf{P}_{\widetilde{X}^n\widetilde{Y}^n}:=\mathsf{P}_{X^nY^n|W_{n-m+1}\wedge\cdots\wedge W_n}$. Lemma 6.4 implies that there exists an index j such that (X,Y) is $(1-\varepsilon)$-embeddable in $(\widetilde{X}^n,\widetilde{Y}^n)$ with $(\widetilde{X}_j,\widetilde{Y}_j)=(X,Y)$ and

$$\varepsilon := 15\sqrt{\frac{1}{n-m}}\sqrt{m\log(|\mathcal{A}||\mathcal{B}|)+\log\Big(\frac{1}{\Pr[W_{n-m+1}\wedge\cdots\wedge W_n]}\Big)}\,.$$

Consider the following strategy for $\mathfrak{G}$. On input (X,Y) Alice and Bob $1-\varepsilon$-embed this into $(\widetilde{X}^n,\widetilde{Y}^n)$ with $(\widetilde{X}_j,\widetilde{Y}_j)=(X,Y)$. Since the resulting distribution has statistical distance at most ε from $\mathsf{P}_{\widetilde{X}^n\widetilde{Y}^n}$, if they output coordinate j of $h_a(\widetilde{X}^n)$ and $h_b(\widetilde{Y}^n)$ they have probability at least $\Pr[W_j|W_{n-m+1}\wedge\cdots\wedge W_n]-\varepsilon$ to win the initial game. The shared randomness can be eliminated (see the remark after Definition 2.2), and thus

$$v(\mathfrak{G}) \ge \Pr[W_j|W_{n-m+1}\wedge\cdots\wedge W_n]-\varepsilon\,.$$ □

7 Parallel repetition theorem

We will need the following recursion in the proof of Theorem 2.5.

Lemma 7.1. *Fix $0 < v < 1$, $c \geq 12$, $\ell \geq 1$, and let $p_0, \ldots, p_n$ be a non-increasing sequence of non-negative reals with $p_0 = 1$ and*

$$p_{m+1} \leq p_m \left(v + \sqrt{\frac{c}{n-m}} \sqrt{m\ell + \log\left(\frac{1}{p_m}\right)} \right). \tag{7.1}$$

Then,

$$p_n \leq \left(1 - \frac{(1-v)^3}{26c}\right)^{\frac{n}{\ell}}. \tag{7.2}$$

Proof. We show by induction that

$$p_m \leq \left(\frac{1+v}{2}\right)^m,$$

as long as $m \leq \frac{(1-v)^2}{12c\ell}(n-m)$. The statement holds for $m = 0$ and we now make a step from m to $m+1$. First, we can assume that

$$p_m \geq \left(\frac{1+v}{2}\right)^{m+1} > \frac{1}{2}^{m+1},$$

as otherwise the induction step is trivial. In this case, (7.1) yields

$$\begin{aligned} p_{m+1} &\leq p_m \cdot \left(v + \sqrt{\frac{c}{n-m}} \sqrt{m\ell + (m+1)} \right) \\ &\leq p_m \cdot \left(v + \sqrt{\frac{1}{n-m}} \sqrt{3cm\ell} \right) \\ &\leq p_m \cdot \left(v + \frac{1-v}{2} \right) = p_m \cdot \frac{1+v}{2}, \end{aligned} \tag{7.3}$$

where we used $m \leq \frac{(1-v)^2}{12c\ell}(n-m)$ in the last inequality.

We get, setting $m^* = \frac{n(1-v)^2}{13c\ell}$ (which satisfies $m^* \leq \frac{(1-v)^2}{12c\ell}(n-m)$ because $c \geq 12$):

$$p_n \leq p_{m^*} \leq \left(\frac{1+v}{2}\right)^{\frac{n(1-v)^2}{13c\ell}}. \tag{7.4}$$

We have

$$\left(\frac{1+v}{2}\right)^{\frac{(1-v)^2}{13c}} = \left(1 - \frac{1-v}{2}\right)^{\frac{(1-v)^2}{13c}} \leq 1 - \frac{(1-v)^3}{26c}, \tag{7.5}$$

where the last inequality follows from $(1-b)^a \leq 1 - ab$ which holds for all $a \in [0,1]$, $b \leq 1$. $\square$

Remark 7.2. The minimal value of the sequence defined by $p_0 := 1$ and the relation

$$p_{m+1} := p_m\big(v + \sqrt{\frac{c}{n-m}}\sqrt{m\ell + \log(1/p_m)}\big)$$

is indeed $\big(1-\Theta((1-v)^3)\big)^{\frac{n}{\ell}}$. The lemma above shows that the minimal value can only be lower. On the other hand, the sequence given by

$$p'_0 := 1\,, \qquad p'_{m+1} := p'_m\Big(v + \sqrt{\frac{m\ell}{n}}\Big)$$

is strictly smaller than the sequence $\{p_j\}_{j\geq 0}$. This sequence does not decrease anymore if $m > m' := n(1-v)^2/\ell$, and

$$p'_{m'} = \prod_{i=0}^{m'-1}\Big(v + \sqrt{\frac{i\ell}{n}}\Big) \geq v^{m'} = (1-(1-v))^{(1-v)^2\frac{n}{\ell}} \approx e^{-(1-v)^3\frac{n}{\ell}} \approx (1-(1-v)^3)^{\frac{n}{\ell}}\,.$$

We now prove the Parallel Repetition Theorem.

Proof of Theorem 2.5. Fix a strategy (h_a, h_b) for $\mathfrak{G}^n$. Then, repeatedly choose the index i_{m+1} for which $\Pr[W_{i_{m+1}} \mid W_{i_1} \wedge \cdots \wedge W_{i_m}]$ is minimized. We set $p_0 := 1$ and $p_m := \Pr[W_{i_1} \wedge \cdots \wedge W_{i_m}]$. Lemma 6.5 implies

$$p_{m+1} \leq p_m \cdot \left(v + 15\sqrt{\frac{1}{n-m}}\sqrt{m\log(|\mathcal{A}||\mathcal{B}|) + \log\Big(\frac{1}{p_m}\Big)}\right). \tag{7.6}$$

We apply Lemma 7.1 and get

$$\Pr[W_1 \wedge \cdots \wedge W_n] = p_n \leq \Big(1 - \frac{(1-v)^3}{6000}\Big)^{\frac{n}{\log(|\mathcal{A}||\mathcal{B}|)}}\,. \tag{7.7}$$

□

8 Improving the rate

Theorem 2.5 shows that n-fold parallel repetition reduces the winning probability from $v(\mathfrak{G})$ to $(1-\Theta(1-v(\mathfrak{G}))^3)^{\frac{n}{\log(|\mathcal{A}||\mathcal{B}|)}}$. As shown in [29], the term $|\mathcal{A}|\cdot|\mathcal{B}|$ in the exponent can be reduced to the the maximum (over x, y) number of (fractional) rectangles needed to cover the 1-entries in $Q(x,y,\cdot,\cdot)$. Here, we show that it can be reduced to a quantity which is possibly smaller in some cases.

Definition 8.1 (Exact Fractional Product Cover)**.** Let $Q : \mathcal{A}\times\mathcal{B} \to \{0,1\}$ be an arbitrary predicate. Two functions $f : \mathcal{A}\times\{1,\ldots,\alpha\} \to [0,1]$ and $g : \mathcal{B}\times\{1,\ldots,\alpha\} \to [0,1]$ form an *exact fractional product cover of size* α for Q if for all a, b we have

$$Q(a,b) = \sum_{i=1}^{\alpha} f(a,i)g(b,i)\,.$$

Clearly, any partition by rectangles gives an exact fractional product cover (by defining $f(a,i)$ and $g(b,i)$ as appropriate predicates). We will prove the following strengthening of Theorem 2.5.

Theorem 8.2. *Let $\mathfrak{G} = (\mathsf{P}_{XY}, Q)$ be a game. Let α be such that for all (x,y) there exists an exact fractional product cover of size α for $Q_{x,y}(a,b) := Q(x,y,a,b)$. If $\alpha > 1$ then*

$$v(\mathfrak{G}^n) \leq \left(1 - \frac{(1-v)^3}{6000}\right)^{\frac{n}{\log(\alpha)}}, \tag{8.1}$$

and if $\alpha = 1$ then

$$v(\mathfrak{G}^n) \leq \left(1 - \frac{(1-v)^2}{6000}\right)^{n}. \tag{8.2}$$

To prove Theorem 8.2 we first need a characterization of fractional product covers by Markov chains.

Lemma 8.3. *Let a distribution $\mathsf{P}_{ABZ} = \mathsf{P}_A \mathsf{P}_B \mathsf{P}_{Z|AB}$ be given for which there exist functions $f(a,z) : \mathcal{A} \times \mathcal{Z} \to [0,1]$ and $g(b,z) : \mathcal{B} \times \mathcal{Z} \to [0,1]$ which satisfy*

$$\mathsf{P}_{Z|A=aB=b}(z) = f(a,z) \cdot g(b,z). \tag{8.3}$$

Then $A \leftrightarrow Z \leftrightarrow B$.

Lemma 8.3 has a converse in the following sense: if $\mathsf{P}_{Z|AB}$ is such that $A \leftrightarrow Z \leftrightarrow B$ for all distributions $\mathsf{P}_A \mathsf{P}_B$, then $\mathsf{P}_{Z|AB}$ is of the form (8.3) for some functions f and g. For completeness, we prove this in Section 10.2.

Proof of Lemma 8.3. We get

$$\begin{aligned}
\mathsf{P}_{A|B=bZ=z}(a) &= \frac{\mathsf{P}_{ABZ}(a,b,z)}{\mathsf{P}_{BZ}(b,z)} \\
&= \frac{\mathsf{P}_A(a)\mathsf{P}_B(b) f(a,z) g(b,z)}{\sum_{a'} \mathsf{P}_B(b)\mathsf{P}_A(a') f(a',z) g(b,z)} \\
&= \frac{\mathsf{P}_A(a) f(a,z)}{\sum_{a'} \mathsf{P}_A(a') f(a',z)} \\
&= \frac{\sum_{b'} \mathsf{P}_A(a)\mathsf{P}_B(b') f(a,z) g(b',z)}{\sum_{a',b'} \mathsf{P}_A(a')\mathsf{P}_B(b') f(a',z) g(b',z)} \\
&= \frac{\mathsf{P}_{AZ}(a,z)}{\mathsf{P}_Z(z)} \\
&= \mathsf{P}_{A|Z=z}(a),
\end{aligned}$$

and thus $\mathsf{P}_{ABZ}(a,b,z) = \mathsf{P}_Z(z)\mathsf{P}_{B|Z=z}(b)\mathsf{P}_{A|B=bZ=z}(a) = \mathsf{P}_Z(z)\mathsf{P}_{B|Z=z}(b)\mathsf{P}_{A|Z=z}(a)$, which means that $A \leftrightarrow Z \leftrightarrow B$. □

Lemma 8.4. *Assume $Q : \mathcal{A} \times \mathcal{B} \to \{0,1\}$ has a fractional product cover of size α. Then there exists a conditional distribution $\mathsf{P}_{Z|AB}$ over[6] $\mathcal{Z} = \{1,\dots,\alpha\} \cup (\mathcal{A} \times \mathcal{B})$ such that*

[6] The intention is that $\{1,\dots,\alpha\} \cap (\mathcal{A} \times \mathcal{B}) = \emptyset$, such that on a given z one can tell whether it's from $\{1,\dots,\alpha\}$ or from $\mathcal{A} \times \mathcal{B}$.

- *for any product distributions* $\mathsf{P}_{AB} = \mathsf{P}_A\mathsf{P}_B$ *and* $\mathsf{P}_{ABZ} = \mathsf{P}_A\mathsf{P}_B\mathsf{P}_{Z|AB}$ *we have* $A \leftrightarrow Z \leftrightarrow B$*;*
- $Q(a,b) = 1 \iff z \in \{1,\dots,\alpha\}$.

Proof. Let f,g be the functions as guaranteed by the exact fractional product cover. We set

$$\mathsf{P}_{Z|A=a,B=b}(z) = \begin{cases} 1 & \text{if } Q(a,b) = 0 \wedge z = (a,b) \\ f(a,z)g(b,z) & \text{if } Q(a,b) = 1 \wedge z \in \{1,\dots,\alpha\} \\ 0 & \text{otherwise.} \end{cases}$$

The properties follow immediately from the definition and Lemma 8.3. □

We strengthen Claim 6.2:

Claim 8.5. *Let* $\mathsf{P}_{X_0Y_0}\mathsf{P}_{X_1Y_1}$ *be a distribution over* $\mathcal{X}_0 \times \mathcal{Y}_0 \times \mathcal{X}_1 \times \mathcal{Y}_1$, $f : \mathcal{X}_0 \times \mathcal{X}_1 \to \mathcal{A}$ *and* $g : \mathcal{Y}_0 \times \mathcal{Y}_1 \to \mathcal{B}$ *be arbitrary. Let* $\mathsf{P}_{Z|AB}$ *be such that for any product distribution* $\mathsf{P}_{AB} = \mathsf{P}_A\mathsf{P}_B$ *the Markov condition* $A \leftrightarrow Z \leftrightarrow B$ *is satisfied. Then, if* Z *is obtained using* $\mathsf{P}_{Z|A=f(X_0,Y_1),B=g(Y_0,Y_1)}$ *we have*

$$X_0X_1 \leftrightarrow X_0ZY_1 \leftrightarrow Y_0Y_1\,. \tag{8.4}$$

Proof. We fix $x_0 \in \mathcal{X}_0$ and $y_1 \in \mathcal{Y}_1$ throughout the proof, and consider everything conditioned on these values. Then $f : \mathcal{X}_1 \to \mathcal{A}$ and $g : \mathcal{Y}_0 \to \mathcal{B}$. We have

$$\mathsf{P}_{X_1Y_0FGZ} = \mathsf{P}_{X_1}\mathsf{P}_{F|X_1}\mathsf{P}_{Y_0}\mathsf{P}_{Y_0|G}\mathsf{P}_{Z|FG} = \mathsf{P}_F\mathsf{P}_{X_1|F}\mathsf{P}_G\mathsf{P}_{Y_0|G}\mathsf{P}_{Z|FG} = \mathsf{P}_Z\mathsf{P}_{F|Z}\mathsf{P}_{G|Z}\mathsf{P}_{X_1|F}\mathsf{P}_{Y_0|G}\,,$$

which implies the claim. □

For the case $\alpha = 1$ we will need a slightly different recursion than the one given in Lemma 7.1.

Lemma 8.6. *Fix* $v < 1$, $c > 5$, *and let* $p_0,\dots,p_n$ *be a sequence of non-increasing non-negative reals with* $p_0 = 1$ *and*

$$p_{m+1} \le p_m\left(v + \sqrt{\frac{c}{n-m}}\sqrt{\log\left(\frac{1}{p_m}\right)}\right). \tag{8.5}$$

Then,

$$p_n \le \left(1 - \frac{(1-v)^2}{10c}\right)^n. \tag{8.6}$$

Proof. We show by induction that $p_m \le (\frac{1+v}{2})^m$ as long as $m+1 \le (n-m)(1-v)/(4c)$. Clearly, this holds for $m = 0$. To make a step from m to $m+1$ we can assume

$$p_m \ge \left(\frac{1+v}{2}\right)^{m+1} = \left(1 - \frac{1-v}{2}\right)^{m+1} \ge \left(1 - \frac{1}{2}\right)^{(1-v)(m+1)} = 2^{-(1-v)(m+1)}$$

(where we used $(1-b)^a \le 1-ab$ for $a \in [0,1]$, $b \le 1$) which means that

$$p_{m+1} \le p_m \cdot \left(v + \sqrt{\frac{c(m+1)(1-v)}{n-m}}\right) \tag{8.7}$$

$$\le p_m \cdot \left(1 + \frac{v}{2}\right) \tag{8.8}$$

as long as $m+1 \le (n-m)(1-v)/(4c)$, which implies the hypothesis. We thus get for $m = n(1-v)/(5c)$

$$p_m \le \left(\frac{1+v}{2}\right)^{\frac{n(1-v)}{5c}}.$$

Finally,

$$\left(\frac{1+v}{2}\right)^{(1-v)/5c} = \left(1 - \frac{(1-v)}{2}\right)^{(1-v)/5c} \le 1 - \frac{(1-v)^2}{10c},$$

again using $(1-b)^a \le 1-ab$. □

Remark 8.7. Lemma 8.6 is essentially tight: in case $p_m \le \left(1 - \Theta((1-v)^2)\right)^n$ equation (8.5) only implies

$$p_{m+1} \le p_m \left(v + c\sqrt{-\log(1-\Theta\big((1-v)^2\big))}\right) = p_m \left(v + c\sqrt{\Theta\big((1-v)^2\big)}\right) = p_m(v+1-v) = p_m\,.$$

We now come to the proof of Theorem 8.2.

Proof (of Theorem 8.2). We first show that Lemma 6.4 still holds if we replace (6.2) by

$$\sum_{j=1}^{k} \varepsilon_j \le 15\sqrt{k}\sqrt{(n-k)\log(\alpha) + \log\Big(\frac{1}{\Pr[W_{k+1} \wedge \cdots \wedge W_n]}\Big)}\,. \tag{8.9}$$

For this, we define the random variables D^k, U^k, $\overline{U}^k$, and T exactly as in the proof of Lemma 6.4. Instead of (6.4) we now define

$$V := (Z_{k+1}, \ldots, Z_n)\,, \tag{8.10}$$

where Z_i is obtained from (A_i, B_i, X_i, Y_i) by a channel that has alphabet size at most α in case W_i, which ensures $A_i \leftrightarrow X_iY_iZ_i \leftrightarrow B_i$ in case A_i and B_i are independent, and for which W_i can be inferred from (X_i, Y_i, Z_i). The existence of such a random variable is ensured by Lemma 8.4 and the fact that for every (x,y) there exists an exact fractional product cover of size α for $Q(x,y,\cdot,\cdot)$.

From Corollary 4.3 we now get

$$\sum_{j=1}^{k} \left\| \mathsf{P}_{TU_jV|W} - \mathsf{P}_{TV|W}\mathsf{P}_{U_j|T} \right\| \le \varepsilon_{\mathrm{Tot}}\,, \tag{8.11}$$

where we set

$$\varepsilon_{\mathrm{Tot}} := \sqrt{k}\sqrt{(n-k)\log(\alpha)+\log\Big(\frac{1}{\Pr[W]}\Big)}\,. \tag{8.12}$$

For a fixed j we define $T^{(\setminus j)}$ as in the proof of Lemma 6.4 and obtain in exactly the same way for $S := (T^{(\setminus j)}, V)$ and

$$\mathsf{P}_{\widetilde{S}\widetilde{X}^n\widetilde{Y}^n} := \mathsf{P}_{SX^nY^n|W} \tag{8.13}$$

the equations

$$\sum_{j=1}^{k}\Big\|\mathsf{P}_{\widetilde{S}\widetilde{X}_j\widetilde{Y}_j} - \mathsf{P}_{XY}\mathsf{P}_{\widetilde{S}|\widetilde{Y}_j}\Big\| \le 3\varepsilon_{\mathrm{Tot}} \tag{8.14}$$

and

$$\sum_{j=1}^{k}\Big\|\mathsf{P}_{\widetilde{S}\widetilde{X}_j\widetilde{Y}_j} - \mathsf{P}_{XY}\mathsf{P}_{\widetilde{S}|\widetilde{X}_j}\Big\| \le 3\varepsilon_{\mathrm{Tot}}\,. \tag{8.15}$$

Again, Corollary 5.3 implies that (X,Y) is $(1-\varepsilon_j)$-embeddable in $(\widetilde{X}_j\widetilde{S}, \widetilde{Y}_j\widetilde{S})$ with $(\widetilde{X}_j, \widetilde{Y}_j) = (X,Y)$ and such that $\sum_{j=1}^{k}\varepsilon_j \le 15\varepsilon_{\mathrm{Tot}}$.

Again we get

$$X^k \leftrightarrow TV \leftrightarrow Y^k\,, \tag{8.16}$$

now using the properties of the Z_i. (This is done as follows: clearly, $X^k \leftrightarrow T \leftrightarrow Y^k$, i. e.for a fixed values t for T the X^k and Y^k are independent. Now, inductively adding Z_i will not change this in any step, due to Claim 8.5.) Claim 6.3 now yields

$$\widetilde{X}^n \leftrightarrow \widetilde{X}_j\widetilde{S} \leftrightarrow \widetilde{Y}^n\widetilde{Y}_j\widetilde{S} \tag{8.17}$$

$$\widetilde{X}^n\widetilde{X}_j\widetilde{S} \leftrightarrow \widetilde{Y}_j\widetilde{S} \leftrightarrow \widetilde{Y}^n\,, \tag{8.18}$$

and Lemma 5.4 completes the proof that (8.9) can replace (6.2) in Lemma 6.4.

From Lemma 6.4 where (6.2) is replaced by (8.9) we apply Lemma 7.1 and get (8.1). To get (8.2) we note first that in this case (8.9) reduces to

$$\sum_{j=1}^{k}\varepsilon_j \le 15\sqrt{k}\sqrt{\log\Big(\frac{1}{\Pr[W_{k+1}\wedge\cdots\wedge W_n]}\Big)}\,, \tag{8.19}$$

and so we can apply Lemma 8.6. □

9 No-signaling strategies

No-signaling strategies are those where the only restriction on the response of Alice and Bob is that they do not *imply* communication.

Definition 9.1 (No-signaling)**.** A pair (h_a, h_b) of functions is *no-signaling* if $h_a : \mathcal{X} \times \mathcal{Y} \times \mathcal{R} \to \mathcal{A}$ and $h_b : \mathcal{X} \times \mathcal{Y} \times \mathcal{R} \to \mathcal{B}$ satisfy

$$\Pr_R[h_a(x,y,R)] = \Pr_R[h_a(x,y',R)]$$
$$\Pr_R[h_b(x,y,R)] = \Pr_R[h_b(x',y,R)]\,,$$

for all x, x', y, y'.

Definition 9.2 (No-signaling value)**.** The no-signaling value $v_{ns}(\mathfrak{G})$ of a *game* $\mathfrak{G} = (\mathsf{P}_{XY}, Q)$ *over* $\mathcal{X} \times \mathcal{Y} \times \mathcal{A} \times \mathcal{B}$ is

$$v_{ns}(\mathfrak{G}) := \max \Pr_{XYR}[Q(X,Y,h_a(X,Y,R),h_b(X,Y,R))]\,,$$

where the maximum is over all no-signaling functions (h_a, h_b).

Clearly, $v(\mathfrak{G}) \le v_{ns}(\mathfrak{G})$, since any local strategy is a no-signaling strategy. We further note that for no-signaling strategies $v_{ns}(\mathfrak{G}^2) > (v_{ns}(\mathfrak{G}))^2$ is also possible, similarly to the local case (see Section 10.1).

We will prove the following upper bound on the no-signaling value of parallel repetition.

Theorem 9.3. *For any game $\mathfrak{G}$ with no-signaling value $v_{ns} := v_{ns}(\mathfrak{G})$ and any integer n we have*

$$v_{ns}(\mathfrak{G}^n) \le \left(1 - \frac{(1-v_{ns})^2}{1000}\right)^n. \tag{9.1}$$

We remark that the proof of this theorem will be much simpler than the proof of Theorem 2.5.

We first show that if P_{XYST} is a distribution which is close to no-signaling (i. e., $\|\mathsf{P}_{XYS} - \mathsf{P}_{XY}\mathsf{P}_{S|X}\| \le \varepsilon$ and $\|\mathsf{P}_{XYT} - \mathsf{P}_{XY}\mathsf{P}_{T|Y}\| \le \varepsilon$) then there exists a no-signaling strategy which produces values (s',t') from (x,y) such that the distribution (S',X,T',Y) is statistically close to the distribution (S,X,T,Y). This will follow by applying the following lemma twice.

Lemma 9.4. *Let P_{S_1T}, P_{S_2} be arbitrary distributions over $\mathcal{S} \times \mathcal{T}$ and $\mathcal{S}$, respectively, where $\mathcal{S}$ and $\mathcal{T}$ are finite. Then there exists a distribution $\mathsf{P}_{\overline{S}\,\overline{T}}$ such that*

$$\|\mathsf{P}_{\overline{S}\,\overline{T}} - \mathsf{P}_{S_1T}\| \le \|\mathsf{P}_{S_1} - \mathsf{P}_{S_2}\| \tag{9.2}$$
$$\|\mathsf{P}_{\overline{S}} - \mathsf{P}_{S_2}\| = 0 \tag{9.3}$$
$$\|\mathsf{P}_{\overline{T}} - \mathsf{P}_{T}\| = 0\,. \tag{9.4}$$

Proof. Let $\mathsf{P}_{\widetilde{S}_1\widetilde{S}_2}$ be the joint distribution guaranteed in Lemma 2.1, (2.4) which satisfies $\mathsf{P}_{\widetilde{S}_1} = \mathsf{P}_{S_1}$, $\mathsf{P}_{\widetilde{S}_2} = \mathsf{P}_{S_2}$ and

$$\Pr[\widetilde{S}_1 = \widetilde{S}_2] = 1 - \|\mathsf{P}_{S_1} - \mathsf{P}_{S_2}\|\,. \tag{9.5}$$

We consider $\mathsf{P}_{\widetilde{S}_1\widetilde{S}_2T} = \mathsf{P}_{\widetilde{S}_1\widetilde{S}_2}\mathsf{P}_{T|S_1}$ and set $\mathsf{P}_{\overline{S}\,\overline{T}} = \mathsf{P}_{\widetilde{S}_2T}$ to be the corresponding marginal. Equations (9.3) and (9.4) are clear. To see (9.2) note that with our definitions $\mathsf{P}_{\overline{S}\,\overline{T}}$ and P_{S_1T} are marginals of the same distribution which additionally satisfies (9.5). □

Lemma 9.5. *Let* $\mathsf{P}_{X_0Y_0}$ *and* P_{XYST} *be arbitrary distributions. If*

$$\|\mathsf{P}_{X_0Y_0}\mathsf{P}_{S|X} - \mathsf{P}_{XYS}\| \le \varepsilon_1\,, \tag{9.6}$$

$$\|\mathsf{P}_{X_0Y_0}\mathsf{P}_{T|Y} - \mathsf{P}_{XYT}\| \le \varepsilon_2\,, \tag{9.7}$$

then there exists a conditional distribution $\mathsf{P}_{S'T'|X'=xY'=y}$ *with* $\mathsf{P}_{S'|X'=xY'=y} = \mathsf{P}_{S'|X'=x}$ *and* $\mathsf{P}_{T'|X'=xY'=y} = \mathsf{P}_{T'|Y'=y}$ *such that*

$$\|\mathsf{P}_{X_0Y_0}\mathsf{P}_{S'T'|XY} - \mathsf{P}_{XYST}\| \le 3\varepsilon_1 + 2\varepsilon_2\,. \tag{9.8}$$

Proof. For fixed x,y we define $\mathsf{P}_{S_0T_0|X=xY=y}$ using Lemma 9.4 with the following properties:

$$\begin{aligned}\|\mathsf{P}_{S_0T_0|X=xY=y} - \mathsf{P}_{ST|X=xY=y}\| &\le \|\mathsf{P}_{S|X=x} - \mathsf{P}_{S|X=xY=y}\|\\ \|\mathsf{P}_{S_0|X=xY=y} - \mathsf{P}_{S|X=x}\| &= 0\\ \|\mathsf{P}_{T_0|X=xY=y} - \mathsf{P}_{T|X=xY=y}\| &= 0\,.\end{aligned}$$

Then, again using Lemma 9.4, we define $\mathsf{P}_{S'T'|X=xY=y}$ such that

$$\begin{aligned}\|\mathsf{P}_{S'T'|X=xY=y} - \mathsf{P}_{S_0T_0|X=xY=y}\| &\le \|\mathsf{P}_{T_0|Y=y} - \mathsf{P}_{T_0|X=xY=y}\|\\ \|\mathsf{P}_{T'|X=xY=y} - \mathsf{P}_{T_0|Y=y}\| &= 0\\ \|\mathsf{P}_{S'|X=xY=y} - \mathsf{P}_{S_0|X=xY=y}\| &= 0\,.\end{aligned}$$

We see that for all pairs x,y we have $\mathsf{P}_{S'|X=xY=y} = \mathsf{P}_{S'|X=x}$ and $\mathsf{P}_{T'|X=xY=y} = \mathsf{P}_{T'|Y=y}$.

We further get

$$\begin{aligned}&\|\mathsf{P}_{X_0Y_0}\mathsf{P}_{S'T'|XY} - \mathsf{P}_{XYST}\|\\ &\quad\le \varepsilon_1 + \|\mathsf{P}_{XY}\mathsf{P}_{S'T'|XY} - \mathsf{P}_{XYST}\|\\ &\quad= \varepsilon_1 + \sum_{x,y,s,t}\big|\mathsf{P}_{XY}(x,y)\mathsf{P}_{S'T'|X=xY=y}(s,t) - \mathsf{P}_{XY}(x,y)\mathsf{P}_{ST|X=xY=y}(s,t)\big|\\ &\quad\le \varepsilon_1 + \sum_{x,y}\mathsf{P}_{XY}(x,y)\Big(\|\mathsf{P}_{S|X=x} - \mathsf{P}_{S|X=xY=y}\| + \|\mathsf{P}_{T|Y=y} - \mathsf{P}_{T|X=xY=y}\|\Big)\\ &\quad\le \varepsilon_1 + \|\mathsf{P}_{XY}\mathsf{P}_{S|X} - \mathsf{P}_{SXY}\| + \|\mathsf{P}_{XY}\mathsf{P}_{T|Y} - \mathsf{P}_{TXY}\|\\ &\quad\le 3\varepsilon_1 + 2\varepsilon_2\,.\end{aligned}$$

□

We can now prove a no-signaling analogue of Lemma 6.5.[7]

Lemma 9.6. *Let a game* $\mathfrak{G} = (Q, \mathsf{P}_{XY})$*, its* n*-fold repetition* $\mathfrak{G}^n$*, and a no-signaling strategy* (h_a, h_b) *for* $\mathfrak{G}^n$ *be given. Let indices* $i_1,\dots,i_m$ *be given. Then there exists an index* i_{m+1} *such that*

$$\begin{aligned}&\Pr[W_{i_{m+1}}|W_{i_1}\wedge\cdots\wedge W_{i_m}]\\ &\qquad\le v_{ns}(\mathfrak{G}) + 10\sqrt{\frac{1}{n-m}}\sqrt{\log\Big(\frac{1}{\Pr[W_{i_1}\wedge\cdots\wedge W_{i_m}]}\Big)}\,.\end{aligned} \tag{9.9}$$

[7] A previous version of the proof of this lemma contained an error, which was first noticed by Oded Regev and Ricky Rosen.

Proof. As in the proof of Lemma 6.5 we assume that $i_\ell = n-\ell+1$ and we define $W := W_{n-m+1} \wedge \cdots \wedge W_n$. The no-signaling property of (h_a, h_b) implies $\mathsf{P}_{X^nY^nA^n} = \mathsf{P}_{X^n}\mathsf{P}_{Y^n|X^n}\mathsf{P}_{A^n|X^n} = \mathsf{P}_{A^nX^n}\mathsf{P}_{Y^n|X^n}$. Thus, when we apply Corollary 4.3 on this distribution (with the event W and the random variables $T = (X^n, A^n)$ and $U_j = Y_j$) we get

$$\sum_{j=1}^{n-m} \left\| \mathsf{P}_{X^nA^nY_j|W} - \mathsf{P}_{X^nA^n|W}\mathsf{P}_{Y_j|X_j} \right\| = \sum_{j=1}^{n-m} \left\| \mathsf{P}_{TY_j|W} - \mathsf{P}_{T|W}\mathsf{P}_{Y_j|T} \right\|$$
$$\leq \sqrt{(n-m)\log\left(\frac{1}{\Pr[W]}\right)}.$$

Taking appropriate marginals this gives

$$\sum_{j=1}^{n-m} \left\| \mathsf{P}_{X_jY_jA_j|W} - \mathsf{P}_{X_jA_j|W}\mathsf{P}_{Y_j|X_j} \right\| \leq \sqrt{(n-m)\log\left(\frac{1}{\Pr[W]}\right)}.$$

Applying Lemma 4.1 once more and rearranging we get

$$\sum_{j=1}^{n-m} \left\| \mathsf{P}_{X_jY_jA_j|W} - \mathsf{P}_{XY}\mathsf{P}_{A_j|X_jW} \right\| \leq 2\sqrt{(n-m)\log\left(\frac{1}{\Pr[W]}\right)}. \tag{9.10}$$

Symmetrically, we obtain

$$\sum_{j=1}^{n-m} \left\| \mathsf{P}_{X_jY_jB_j|W} - \mathsf{P}_{XY}\mathsf{P}_{B_j|Y_jW} \right\| \leq 2\sqrt{(n-m)\log\left(\frac{1}{\Pr[W]}\right)}. \tag{9.11}$$

From (9.10), (9.11), and Lemma 9.5 we get that there exists a distribution $\mathsf{P}_{A'_jB'_j|XY}$ which can be implemented by no-signaling functions and for which

$$\sum_{j=1}^{n-m} \left\| \mathsf{P}_{XY}\mathsf{P}_{A'_jB'_j|XY} - \mathsf{P}_{X_jY_jA_jB_j|W} \right\| \leq 10\sqrt{(n-m)\left(\log\left(\frac{1}{\Pr[W]}\right)\right)}.$$

Thus, if Alice and Bob use the strategy implied by $\mathsf{P}_{A'_jB'_j|XY}$ (which is no-signaling) they can win the initial game with probability $\Pr[W_j|W] - 10\sqrt{1/(n-m)}\sqrt{\log(1/\Pr[W])}$ for some j, which implies the lemma. □

Proof (of Theorem 9.3). Fix a no-signaling strategy (h_a, h_b) for $\mathfrak{G}$. As in the proof of Theorem 2.5 we repeatedly select indices i_{m+1} such that $\Pr[W_{i_{m+1}}|W_{i_1} \wedge \cdots \wedge W_{i_m}]$ is minimized. Let $p_m := \Pr[W_{i_1} \wedge \cdots \wedge W_{i_m}]$. Lemma 9.6 implies

$$p_{m+1} \leq p_m \cdot \left(v + 10\sqrt{\frac{1}{n-m}\log\left(\frac{1}{p_m}\right)}\right). \tag{9.12}$$

From (9.12) we apply Lemma 8.6 to get (9.1). □

10 Appendix

10.1 Non-triviality

Local case We quickly reproduce a slight modification[8] of Fortnow's example [18] which shows that $v(\mathfrak{G}^2) > (v(\mathfrak{G}))^2$ is possible. The same variation was also considered by Feige and Lovász [16].

The game we describe is over bits (i. e., all the queries and all the responses are bits). We set

$$\mathsf{P}_{XY}(0,0) := \mathsf{P}_{XY}(0,1) := \mathsf{P}_{XY}(1,0) := \frac{1}{3},$$

and define

$$Q(x,y,a,b) := \big((x \vee a) \neq (y \vee b)\big). \tag{10.1}$$

This can be described in words: Alice and Bob each receive a bit, and at least one of these bits is 0. If both players receive 0, exactly one player must respond with 1. If one of the players receives 1, the other must respond with 0.

We first show that for this game $v = \frac{2}{3}$. Clearly, $v \geq \frac{2}{3}$ (e. g., both players always answer 0). To show $v \leq \frac{2}{3}$ we check all deterministic strategies. If both players reply 0 on query 0, this fails in case $x = y = 0$ (and thus with probability $\frac{1}{3}$). If one player, w.l.o.g. Alice, answers 0 with 1 the players fail in case $x = 0$ and $y = 1$.

If this game is repeated twice in parallel, setting $(a_1,a_2) := (x_2,x_1)$, $(b_1,b_2) := (y_2,y_1)$ also wins with probability $\frac{2}{3}$. One can check this as follows: for every fixed query (x_1,y_1) answering with (x_2,y_2) wins the first subgame with probability $\frac{2}{3}$ (one checks all 3 cases). Moreover, with this strategy

$$Q(x_1,y_1,a_1,b_1) \equiv Q(x_2,y_2,a_2,b_2)$$

which implies the claim.

No-signaling case We now show that for the above game

$$v(\mathfrak{G}) = v(\mathfrak{G}^2) = v_{ns}(\mathfrak{G}) = v_{ns}(\mathfrak{G}^2). \tag{10.2}$$

Previously, it was known that quantum strategies do not help Alice and Bob to win this game [35] (in either the single instance case or where two parallel instances are used).

To show (10.2) it is sufficient to show that $v(\mathfrak{G}) \geq v_{ns}(\mathfrak{G})$ (since $v_{ns}(\mathfrak{G}) \geq v_{ns}(\mathfrak{G}^2) \geq v(\mathfrak{G}^2) = v(\mathfrak{G})$ is already known). There are two ways to see this. First, one can notice that the joint probability of Alice's and Bob's reply only matters if $x = y = 0$; i. e., only for one query. In such a case one can always get a local strategy which is as good as a given no-signaling strategy. Alternatively, let p be the probability that Alice replies 0 on query 0 and q be the probability that Bob replies 0 on query 0. In this case, the players win with probability at most p on query $(x,y) = (0,1)$, with probability at most q on query $(1,0)$, and with probability at most $(1-p)+(1-q)$ on query $(0,0)$, which gives an overall winning probability of at most $\frac{2}{3}$.

[8] Fortnow also lets the referee choose $x = y = 1$ with some probability, in which case the players cannot win the game.

10.2 Converse of Lemma 8.3

We show here that Lemma 8.3 can be strengthened to get an "if and only if" condition.

Lemma 10.1. *Let a conditional distribution $\mathsf{P}_{Z|AB}$ over finite sets $\mathcal{A}$, $\mathcal{B}$, $\mathcal{Z}$ be given. If for all product distributions $\mathsf{P}_{AB} = \mathsf{P}_A\mathsf{P}_B$ the Markov condition $A \leftrightarrow Z \leftrightarrow B$ is satisfied then there exist functions $f(a,z) : \mathcal{A} \times \mathcal{Z} \to [0,1]$ and $g(b,z) : \mathcal{B} \times \mathcal{Z} \to [0,1]$ such that*

$$\mathsf{P}_{Z|A=aB=b}(z) = f(a,z) \cdot g(b,z)\,. \tag{10.3}$$

Proof. Fix an arbitrary z throughout the proof, and consider arbitrary elements $a,a' \in \mathcal{A}$ and $b,b' \in \mathcal{B}$. We set $\mathsf{P}_A(a) = \mathsf{P}_A(a') = \frac{1}{2}$ and $\mathsf{P}_B(b) = \mathsf{P}_B(b') = \frac{1}{2}$. The Markov condition implies

$$\mathsf{P}_{A|Z=zB=b}(a) = \mathsf{P}_{A|Z=zB=b'}(a)$$

which is equivalent to

$$\frac{\mathsf{P}_{ABZ}(a,b,z)}{\mathsf{P}_{ABZ}(a,b,z) + \mathsf{P}_{ABZ}(a',b,z)} = \frac{\mathsf{P}_{ABZ}(a,b',z)}{\mathsf{P}_{ABZ}(a,b',z) + \mathsf{P}_{ABZ}(a',b',z)}$$

or (because of our choice of P_{AB})

$$\frac{\mathsf{P}_{Z|A=a,B=b}(z)}{\mathsf{P}_{Z|A=a,B=b}(z) + \mathsf{P}_{Z|A=a',B=b}(z)} = \frac{\mathsf{P}_{Z|A=a,B=b'}(z)}{\mathsf{P}_{Z|A=a,B=b'}(z) + \mathsf{P}_{Z|A=a',B=b'}(z)}\,.$$

Analogously one gets (by swapping the roles of a and a')

$$\frac{\mathsf{P}_{Z|A=a',B=b}(z)}{\mathsf{P}_{Z|A=a,B=b}(z) + \mathsf{P}_{Z|A=a',B=b}(z)} = \frac{\mathsf{P}_{Z|A=a',B=b'}(z)}{\mathsf{P}_{Z|A=a,B=b'}(z) + \mathsf{P}_{Z|A=a',B=b'}(z)}\,.$$

Together, this implies

$$\mathsf{P}_{Z|A=a,B=b}(z)\mathsf{P}_{Z|A=a',B=b'}(z) = \mathsf{P}_{Z|A=a,B=b'}(z)\mathsf{P}_{Z|A=a',B=b}(z)\,. \tag{10.4}$$

Fix a^* with $\sum_{b\in\mathcal{B}} \mathsf{P}_{Z|A=a^*,B=b}(z) > 0$ (if there is no such a^* we let f and g be the zero functions) and let b^* be such that $\mathsf{P}_{Z|A=a^*,B=b^*}(z) \geq \mathsf{P}_{Z|A=a^*,B=b}(z)$ for all b (note that $\mathsf{P}_{Z|A=a^*,B=b^*}(z) > 0$). We define

$$f(a,z) := \mathsf{P}_{Z|A=a,B=b^*}(z)\,,$$
$$g(b,z) := \frac{\mathsf{P}_{Z|A=a^*,B=b}(z)}{\mathsf{P}_{Z|A=a^*,B=b^*}(z)}\,,$$

and we note that because of our choice of a^* and b^* both $f,g \in [0,1]$. We further get, for any a, b:

$$f(a,z)g(b,z) = \frac{\mathsf{P}_{Z|A=a,B=b^*}(z)\mathsf{P}_{Z|A=a^*,B=b}(z)}{\mathsf{P}_{Z|A=a^*,B=b^*}(z)} = \mathsf{P}_{Z|A=a,B=b}(z)\,,$$

where the last equation follows from (10.4). □

Acknowledgments

I would like to thank Paul Beame, Georges Baatz, Ryan O'Donnell, Johan Håstad, Bartosz Przydatek, Anup Rao, Ran Raz, Renato Renner, Ricky Rosen, Stefano Tessaro, Stefan Wolf, and Jürg Wullschleger for helpful discussions. Paul Beame pointed out [29] to me and explained how the strengthening given there can be adapted to the proof given here; he also simplified an argument in a previous version of this paper. Anup Rao simplified an argument in the proof of Lemma 5.2. Oded Regev and Ricky Rosen found an error in a previous version of the proof of Lemma 9.6. The anonymous referees gave helpful comments. I was supported by the Swiss National Science Foundation, project no. 200020-103847/1.

References

[1] SANJEEV ARORA, CARSTEN LUND, RAJEEV MOTWANI, MADHU SUDAN, AND MARIO SZEGEDY: Proof verification and hardness of approximation problems. In *Proc. 33rd FOCS*, pp. 14–23. IEEE Comp. Soc. Press, 1992. [doi:10.1109/SFCS.1992.267823]. 142

[2] SANJEEV ARORA AND SHMUEL SAFRA: Probabilistic checking of proofs; a new characterization of NP. In *Proc. 33rd FOCS*, pp. 2–13. IEEE Comp. Soc. Press, 1992. [doi:10.1109/SFCS.1992.267824]. 142

[3] BOAZ BARAK, MORITZ HARDT, ISHAY HAVIV, ANUP RAO, ODED REGEV, AND DAVID STEURER: Rounding parallel repetition of unique games. In *Proc. 49th FOCS*, pp. 374–383. IEEE Comp. Soc. Press, 2008. [doi:10.1109/FOCS.2008.31]. 144

[4] PAUL BEAME: Personal communication, 2006. 146

[5] MICHAEL BEN-OR, SHAFI GOLDWASSER, JOE KILIAN, AND AVI WIGDERSON: Multi-prover interactive proofs: How to remove intractability assumptions. In *Proc. 20th STOC*, pp. 113–132. ACM Press, 1988. [doi:10.1145/62212.62223]. 142

[6] GILLES BRASSARD, ANNE BROADBENT, AND ALAIN TAPP: Quantum pseudo-telepathy. *Foundations of Physics*, 35(11):1877–1907, 2005. [doi:10.1007/s10701-005-7353-4, arXiv:quant-ph/0407221]. 143

[7] ANDREI Z. BRODER, STEVEN C. GLASSMAN, MARK S. MANASSE, AND GEOFFREY ZWEIG: Syntactic clustering of the web. *Computer Networks*, 29(8–13):1157–1166, 1997. 143

[8] JIN-YI CAI, ANNE CONDON, AND RICHARD J. LIPTON: On games of incomplete information. *Theoretical Computer Science*, 103(1):25–38, 1992. [doi:10.1016/0304-3975(92)90085-T]. 142

[9] BORIS S. CIREL'SON: Quantum generalizations of Bell's inequality. *Lett. Math. Phys.*, 4(2):93–100, 1980. [doi:10.1007/BF00417500]. 142

[10] JOHN F. CLAUSER, MICHAEL A. HORNE, ABNER SHIMONY, AND RICHARD A. HOLT: Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.*, 23(15):880–884, 1969. [doi:10.1103/PhysRevLett.24.549]. 142

[11] RICHARD CLEVE, WILLIAM SLOFSTRA, FALK UNGER, AND SARVAGYA UPADHYAY: Perfect parallel repetition theorem for quantum XOR proof systems. *Computational Complexity*, 17(2):282–299, 2008. [doi:10.1007/s00037-008-0250-4]. 143

[12] THOMAS M. COVER AND JOY A. THOMAS: *Elements of Information Theory*. John Wiley & Sons, Inc., second edition, 1991. ISBN 0-471-24195-4. 145, 149

[13] URIEL FEIGE: On the success probability of the two provers in one-round proof systems. In *Proc. 6th Ann. Structure in Complexity Theory Conf.*, pp. 116–123. IEEE Comp. Soc. Press, 1991. [doi:10.1109/SCT.1991.160251]. 142

[14] URIEL FEIGE, SHAFI GOLDWASSER, LÁSZLÓ LOVÁSZ, SHMUEL SAFRA, AND MARIO SZEGEDY: Approximating clique is almost NP-complete (preliminary version). In *Proc. 32nd FOCS*, pp. 2–12. IEEE Comp. Soc. Press, 1991. [doi:10.1109/SFCS.1991.185341]. 142

[15] URIEL FEIGE, GUY KINDLER, AND RYAN O'DONNELL: Understanding parallel repetition requires understanding foams. In *Proc. IEEE Conf. Comput. Complexity*, pp. 179–192. IEEE Comp. Soc. Press, 2007. [doi:10.1109/CCC.2007.39]. 144

[16] URIEL FEIGE AND LÁSZLÓ LOVÁSZ: Two-prover one-round proof systems: Their power and their problems. In *Proc. 24th STOC*, pp. 733–744. ACM Press, 1992. [doi:10.1145/129712.129783]. 142, 167

[17] URIEL FEIGE AND OLEG VERBITSKY: Error reduction by parallel repetition – a negative result. *Combinatorica*, 22(4):461–478, 2002. [doi:10.1007/s00493-002-0001-0]. 142

[18] LANCE FORTNOW: *Complexity-Theoretic Aspects of Interactive Proof Systems*. PhD thesis, Massachusetts Institute of Technology, 1989. 142, 167

[19] LANCE FORTNOW, JOHN ROMPEL, AND MICHAEL SIPSER: On the power of multi-prover interactive proof systems. *Theoretical Computer Science*, 134(2):545–557, 1994. [doi:10.1016/0304-3975(94)90251-8]. 142

[20] SREENIVAS GOLLAPUDI AND RINA PANIGRAHY: A dictionary for approximate string search and longest prefix search. In *Proc. of 15th ACM Intern. Conf. on Inf. and Knowl. Manag. (CIKM)*, pp. 768–775. ACM Press, 2006. [doi:10.1145/1183614.1183723]. 143

[21] THOMAS HOLENSTEIN: Parallel repetition: Simplifications and the no-signaling case. In *Proc. 39th STOC*, pp. 411–419. ACM Press, 2007. [doi:10.1145/1250790.1250852]. 141, 143

[22] THOMAS HOLENSTEIN AND RENATO RENNER: On the randomness of independent experiments, 2006. [arXiv:cs.IT/0608007]. 148

[23] SUBHASH KHOT: On the power of unique 2-prover 1-round games. In *Proc. 34th STOC*, pp. 767–775. ACM Press, 2002. [doi:10.1145/509907.510017]. 144

[24] JON M. KLEINBERG AND ÉVA TARDOS: Approximation algorithms for classification problems with pairwise relationships: Metric labeling and Markov random fields. *J. ACM*, 49(5):616–639, 2002. [doi:10.1145/585265.585268]. 143

[25] DROR LAPIDOT AND ADI SHAMIR: A one-round, two-prover, zero-knowledge protocol for NP. *Combinatorica*, 15(2):203–214, 1995. [doi:10.1007/BF01200756]. 142

[26] MARK MANASSE, FRANK MCSHERRY, AND KUNAL TALWAR: Consistent weighted sampling. Manuscript, 2007. 143

[27] UDI MANBER: Finding similar files in a large file system. In *USENIX Winter*, pp. 1–10, 1994. 143

[28] MICHAEL A. NIELSEN AND ISAAC L. CHUANG: *Quantum Computation and Quantum Information*. Cambridge University Press, 2000. ISBN 0-521-63503-9. 142

[29] ITZHAK PARNAFES, RAN RAZ, AND AVI WIGDERSON: Direct product results and the GCD problem, in old and new communication models. In *Proc. 29th STOC*, pp. 363–372. ACM Press, 1997. [doi:10.1145/258533.258620]. 142, 143, 146, 159, 169

[30] SANDU POPESCU AND DANIEL ROHRLICH: Nonlocality as an axiom for quantum theory. *Foundations of Physics*, 24(3):379–385, March 1994. [doi:10.1007/BF02058098, arXiv:quant-ph/9508009]. 142

[31] ANUP RAO: Parallel repetition in projection games and a concentration bound. In *Proc. 40th STOC*, pp. 1–10. ACM Press, 2008. [doi:10.1145/1374376.1374378]. 143, 144, 146

[32] RAN RAZ: A parallel repetition theorem. *SIAM J. Comput.*, 27(3):763–803, 1998. [doi:10.1137/S0097539795280895]. 142, 143, 148

[33] RAN RAZ: A counterexample to strong parallel repetition. In *Proc. 49th FOCS*, pp. 369–373. IEEE Comp. Soc. Press, 2008. [doi:10.1109/FOCS.2008.49]. 144

[34] OLEG VERBITSKY: Towards the parallel repetition conjecture. In *Proc. 9th Ann. Structure in Complexity Theory Conf.*, pp. 304–307. IEEE Comp. Soc. Press, 1994. [doi:10.1109/SCT.1994.315794]. 142

[35] JOHN WATROUS: A note on parallel repetition of two-prover quantum-interactive proofs, 2002. Manuscript. 167

AUTHOR

Thomas Holenstein
postdoctoral researcher
Princeton University
tholenst@princeton.edu
http://www.cs.princeton.edu/~tholenst

ABOUT THE AUTHOR

THOMAS HOLENSTEIN obtained his Ph. D. from ETH Zürich in 2006. His advisor was Ueli Maurer. His CS interests include cryptography, interactive proofs, semidefinite programming, and information theory. He enjoys playing Go and riding his bicycle to work.