SIAM J. COMPUT.
Vol. 27, No. 3, pp. 763–803, June 1998

011

A PARALLEL REPETITION THEOREM[*]

RAN RAZ[†]

Abstract. We show that a parallel repetition of any two-prover one-round proof system (MIP$(2,1)$) decreases the probability of error at an exponential rate. No constructive bound was previously known. The constant in the exponent (in our analysis) depends only on the original probability of error and on the total number of possible answers of the two provers. The dependency on the total number of possible answers is logarithmic, which was recently proved to be almost the best possible [U. Feige and O. Verbitsky, *Proc. 11th Annual IEEE Conference on Computational Complexity*, IEEE Computer Society Press, Los Alamitos, CA, 1996, pp. 70–76].

Key words. interactive proofs, parallel repetition, direct product

AMS subject classifications. 68Q25, 68R99

PII. S0097539795280895

1. Introduction and basic notations.

1.1. History and motivation.

We consider two-prover one-round proof systems (MIP$(2,1)$), as introduced in [12]. In an MIP$(2,1)$ proof system, two computationally unbounded provers (that are not allowed to communicate with each other) try to convince a probabilistic polynomial time verifier that a common input I belongs to a prespecified language L. The proof proceeds in one round: The verifier generates a pair of questions (x, y), based on I and on a random string r, and sends x to the first prover and y to the second prover. The first prover responds by sending u, and the second prover responds by sending v. Based on I, r, x, y, u, v, the verifier decides whether to accept or reject the conjecture: $I \in L$. In what follows, the strategy of the verifier (the way to generate the questions and to decide whether $I \in L$) is called "a proof system," and the pair of strategies of the provers is called "a protocol." A language L is in MIP$(2,1)$ with probability of error ϵ if there exists a proof system (of this type), such that: (1) For $I \in L$, there exists a protocol that causes the verifier to always accept; (2) For $I \notin L$, for any possible protocol, the verifier accepts with probability smaller than ϵ. For the exact definitions of MIP$(2,1)$ proof systems and the family of languages MIP$(2,1)$, see [12, 23].

The class of languages MIP$(2,1)$ turned out to be very powerful. In particular, it follows from [7, 37, 25] that NEXPTIME = MIP$(2,1)$ with exponentially small probability of error. MIP$(2,1)$ proof systems have cryptographic applications (see [12, 13, 36, 18]), and have also been used as a starting point to prove that certain optimization problems are hard to approximate (see [22, 4, 5, 25, 6, 10, 38, 8]). For more discussion about the class MIP$(2,1)$ and its applications, see [23] and the references there.

Sequential repetition of MIP$(2,1)$ proof systems decreases the probability of error exponentially, but requires multiple rounds. Parallel repetition preserves the number of rounds. At what rate is the probability of error decreased by parallel repetition? At first, it was believed that, as in the sequential case, repeating a proof system k times

[*] Received by the editors January 17, 1995; accepted for publication (in revised form) July 29, 1996. A preliminary version of the paper appeared in Proceeding of the 27th Annual ACM Symposium on the Theory of Computing, ACM, New York, 1995, pp. 447–556.

http://www.siam.org/journals/sicomp/27-3/28089.html

[†] Department of Applied Mathematics, Weizmann Institute, Rehovot 76100, Israel (ranraz@wisdom.weizmann.ac.il).

in parallel decreases the probability of error to ϵ^k (see [26, 28]). A counter example for this conjecture was given in [21] (see also [36, 19, 25, 27]). For years it was not even known whether parallel repetition can make the probability of error arbitrarily small. This was recently proved in [44]. No constructive bound, however, was given there for the number of repetitions required to decrease the probability of error below a given bound.

In the simpler special case, where the provers' questions, x and y, are chosen independently, it has been known that repeating a proof system k times in parallel decreases the probability of error exponentially. The proof was first given in [15], and the bound was further improved in [36, 39, 19, 1]. Another special case, a tree-like-measure case, was recently settled in [45]. The constant in the exponent in all those proofs, however, is (at least) polynomial in the number of possible questions (x, y). We remark that in most interesting cases, the questions x, y are not chosen independently and the measure is not tree-like.

Researchers were interested in analyzing the probability of error of a parallel repetition, not only as a mathematical problem, but also because an efficient technique to decrease the probability of error was needed. In the literature there are many results that use different techniques to decrease the probability of error [19, 33, 37, 25, 8, 23, 43]. For certain applications, however, these techniques are insufficient. Parallel repetition was suggested as a technique to decrease the probability of error, because it was believed to be very efficient, and because it preserves many canonical properties of the proof system (e.g., zero knowledge).

Since the appearance of a preliminary version of this paper in STOC95 [40], our results were used by [9, 32] to improve many of the hardness results for approximation of optimization problems (see also [2, 3, 24, 11, 30, 31]). In particular, [32] uses our results to prove very impressive optimal inapproximability results for some of the most basic optimization problems (e.g., for Max-3-SAT). Our result was also used by [42] to prove a new direct sum theorem for probabilistic communication complexity.

The reader can find more about the problem of parallel repetition in [19, 25, 23, 20].

1.2. Main result.

We will concentrate on the deterministic case, where the verifier decides whether to accept or reject the conjecture: $I \in L$, based on I, x, y, u, v only (and not on the random string r). The probabilistic case, where this decision depends also on r, is shortly discussed in section 7.

Following [15, 19, 25], we model the problem as a problem on games:
A game G consists of four finite sets X, Y, U, V, with a probability measure $\mu : X \times Y \to \mathbf{R}^+$, and a predicate $Q : X \times Y \times U \times V \to \{0, 1\}$. We think of Q also as a set. A protocol for G consists of a function $u : X \to U$, and a function $v : Y \to V$. The value of the protocol is defined as

$$\sum_{X \times Y} \mu(x, y) Q(x, y, u(x), v(y)),$$

i.e., the μ-probability that $Q(x, y, u(x), v(y)) = 1$. The value of the game, $w(G)$, is defined to be the maximal value of all protocols for G. The answer size of the game, $s(G)$, is defined by $s(G) = |U||V|$.

We think of G as a game for two players (provers): Player I receives $x \in X$, and player II receives $y \in Y$, according to the pair distribution $\mu(x, y)$. We think of x, y as the inputs for the game. Each player doesn't know the other player's input. Player I has to give an "answer" $u' \in U$, and player II has to give an "answer" $v' \in V$.

The goal of the players is to maximize the μ-probability that Q is satisfied (i.e., the probability that $Q(x, y, u', v') = 1$). We remark that this probability corresponds to the probability of error.

The game $G \otimes G$ consists of the sets $X \times X$, $Y \times Y$, $U \times U$, $V \times V$, with the measure

$$\mu \otimes \mu((x_1, x_2), (y_1, y_2)) = \mu(x_1, y_1)\mu(x_2, y_2)$$

and the predicate

$$Q \otimes Q((x_1, x_2), (y_1, y_2), (u_1, u_2), (v_1, v_2)) = Q(x_1, y_1, u_1, v_1)Q(x_2, y_2, u_2, v_2).$$

In the same way, the game $G^{\otimes k}$ consists of the sets X^k, Y^k, U^k, V^k, with the measure

$$\mu^{\otimes k}(x, y) = \prod_{i=1}^{k} \mu(x_i, y_i)$$

and the predicate

$$Q^{\otimes k}(x, y, u, v) = \prod_{i=1}^{k} Q(x_i, y_i, u_i, v_i),$$

where $x \in X^k, y \in Y^k, u \in U^k, v \in V^k$. Here and throughout, z_i stands for the ith coordinate of a vector z. We denote $G^{\otimes k}, Q^{\otimes k}, X^k, Y^k, U^k, V^k, \mu^{\otimes k}$ also by $\bar{G}, \bar{Q}, \bar{X}, \bar{Y}, \bar{U}, \bar{V}, \bar{\mu}$.

Assume w.l.o.g. that $s(G) \geq 2$. In this paper we prove the following parallel repetition theorem.

THEOREM 1.1. *There exists a global function $W : [0, 1] \to [0, 1]$, with $z < 1 \Rightarrow W(z) < 1$, such that given a game G, with value $w(G)$, and answer size $s(G) \geq 2$:*

$$w(G^{\otimes k}) \leq W(w(G))^{k/\log_2(s(G))}.$$

The exact behavior of the function $W(z)$ is not the focus of this paper. We will make, however, a few comments about the function $W(z)$ implicit in the paper:

1. When z tends to 0, the function $W(z)$ (implicit in this paper) tends to a constant $const_1 > 0$. In fact, for every $0 < z \leq 1$, $W(z) > const_1$. It is still not clear whether a tendency of $W(z)$ to 0, when z tends to 0, can be achieved.
2. Obviously, if $w(G)$ is a global constant (e.g., $w(G) = 1/2$) then $W(w(G))$ is just a different global constant. For example, there exists a constant $const_2 < 1$, s.t. for every $0.01 \leq z \leq 0.99$, $const_1 < W(z) < const_2$.
3. Obviously, when z tends to 1, $W(z)$ also tends to 1. It will be simpler to denote $t = 1 - z$, and to analyze the behavior of $[1 - W(1 - t)]$ when t tends to 0. It is implicit in this paper that there exists a (small) constant $const_3$, (1/34 seems to be enough), s.t. when t tends to 0, $[1 - W(1 - t)]$ can be bounded by $O(t^{const_3})$. For example, there exists a constant $const_4$, s.t. for every $t \leq 0.01$, $[1 - W(1 - t)] \leq const_4 \cdot t^{const_3}$, i.e.,

$$W(1 - t) \geq 1 - const_4 \cdot (1 - z)^{const_3}.$$

It was recently shown in [27] that in certain examples the number of repetitions, k, required to decrease the probability of error from $w(G) = 1/2$ to $w(G^{\otimes k}) \leq 1/8$ is

$$\Omega\left(\frac{\log_2(s(G))}{\log_2 \log_2(s(G))}\right).$$

This shows that the factor $\log_2(s(G))$ in Theorem 1.1 is almost the best possible.

It was observed by [42] that in Theorem 1.1, the term $\log_2(s(G))$ can be replaced with the (possibly smaller) $CC(G)$, or with the (possibly smaller) $\rho(G)$, defined in the following way.

For every $(x,y) \in X \times Y$, define $Q_{x,y} : U \times V \to \{0,1\}$, by $Q_{x,y}(u,v) = Q(x,y,u,v)$. Define $CC_{x,y}$ to be the deterministic communication complexity of the function $Q_{x,y}$, and define $\rho_{x,y}$ to be the exact cover number of the same function (for the definitions see [42, 34]). Then define $CC(G)$ to be the maximum, taken over x,y, of $CC_{x,y}$, and define $\rho(G)$ to be the maximum, taken over x,y, of $\rho_{x,y}$. It is well known (and very easy to prove) that

$$\rho(G) \leq CC(G) \leq \log_2(s(G)).$$

Throughout the paper (sections 2, 3, 4, 5, 6), X,Y,U,V,Q,μ refer to the game G, from Theorem 1.1. Similarly, $k,\bar{G},\bar{X},\bar{Y},\bar{U},\bar{V},\bar{Q},\bar{\mu}$ refer to Theorem 1.1 as well. Denote, for the rest of the paper $s = s(G) = |U||V|$.

1.3. Main technical theorem.

In what follows, we will sometimes denote the game G by G_μ, and denote its value by $w(\mu)$. We will sometimes say that the protocol is a protocol for μ (as X,Y,U,V,Q will be fixed). We will look at the values $w(\gamma)$, for different probability measures $\gamma : X \times Y \to \mathbf{R}^+$.

For a measure $\alpha : \Omega \to \mathbf{R}^+$, and a set $C \subset \Omega$, we use the usual notation

$$\alpha(C) = \sum_{z \in C} \alpha(z).$$

For a probability measure $\alpha : \Omega \to \mathbf{R}^+$, and a set $C \subset \Omega$, denote by $\alpha_C : \Omega \to \mathbf{R}^+$ the probability measure

$$\alpha_C(z) = \begin{cases} 0 & \text{for } z \notin C, \\ \frac{\alpha(z)}{\alpha(C)} & \text{for } z \in C. \end{cases}$$

We will think of α_C also as $\alpha_C : C \to \mathbf{R}^+$. This definition makes sense only if $\alpha(C) > 0$. For C, with $\alpha(C) = 0$, define $\alpha_C(z)$ to be identically 0.

More generally, the term $\frac{0}{0}$ can appear in some places in this paper. Unless said otherwise, $\frac{0}{0}$ is defined to be 0. The term $0z$ is defined to be 0, even if $z = \infty$, or z is undefined. This can occur when we take the expectancy (or a weighted average) of a variable z, which is undefined with probability 0 (or with a weight of 0).

For a set $C \subset \bar{X} \times \bar{Y}$, define the game $G_C^{\otimes k} = \bar{G}_C$ to consist of $\bar{X},\bar{Y},\bar{U},\bar{V}$, with the measure $\bar{\mu}_C$ and the predicate $\bar{Q}$. We think of $\bar{G}_C$ as the restriction of $\bar{G}$ to the set C.

In the proof we will always work with a product set $A = A_X \times A_Y$, where $A_X \subset \bar{X}$, $A_Y \subset \bar{Y}$. Since in most parts of the paper we work with one specific $A = A_X \times A_Y$, it will be convenient to denote (throughout the paper) the measure

$\bar{\mu}_A$ by $\bar{\pi}$. The game $\bar{G}_A$ will also be denoted by $\bar{G}_{\bar{\pi}}$, and its value will also be denoted by $\bar{w}(\bar{\pi})$.

Define the predicate $\bar{Q}^i : \bar{X} \times \bar{Y} \times U \times V \to \{0,1\}$ by

$$\forall x \in \bar{X}, y \in \bar{Y}, u' \in U, v' \in V \;:\; \bar{Q}^i(x,y,u',v') = Q(x_i, y_i, u', v').$$

Define the game $\bar{G}^i_A$ to consist of $\bar{X}, \bar{Y}, U, V$ with the measure $\bar{\pi}$, and the predicate $\bar{Q}^i$. We also denote $\bar{G}^i_A$ by $\bar{G}^i_{\bar{\pi}}$, and its value by $w_i(\bar{\pi})$. We think of $\bar{G}^i_{\bar{\pi}}$ as the restriction of $\bar{G}_{\bar{\pi}}$ to one coordinate. Notice that for an input $(x,y) \in \bar{X} \times \bar{Y}$, the game $\bar{G}_{\bar{\pi}}$ just means playing simultaneously all the games $\bar{G}^i_{\bar{\pi}}$ (on the same input).

The following is our main technical theorem. The theorem claims that if $A = A_X \times A_Y$ is large, then $w_i(\bar{\pi})$ is small for at least one coordinate i.

THEOREM 1.2. *There exists a global function* $W_2 : [0,1] \to [0,1]$*, with* $z < 1 \Rightarrow W_2(z) < 1$*, and a global constant* c_0*, such that for all games* G*: for any* k*, and any product set* $A = A_X \times A_Y \subset \bar{X} \times \bar{Y}$ *(where* $A_X \subset \bar{X}, A_Y \subset \bar{Y}$*), and any* $0 \le \Delta \le 1$*, if* $-\log_2 \bar{\mu}(A) \le \Delta k$ *(i.e.,* $\bar{\mu}(A) \ge 2^{-\Delta k}$*) then there exists* i *with*

$$w_i(\bar{\pi}) \le W_2(w(G)) + c_0 \Delta^{1/16}.$$

Again, achieving the best function W_2 and the best constant c_0, and improving the constant 1/16, are not the focus of this paper.

1.4. More notations and basic facts. For a measure or function $\alpha : \Omega^k \to \mathbf{R}$ define α^i to be the projection of α on the ith coordinate. Thus, for $a \in \Omega$

$$\alpha^i(a) = \sum_{\{z \in \Omega^k \mid z_i = a\}} \alpha(z).$$

In particular for $\alpha : \bar{X} \times \bar{Y} \to \mathbf{R}$, define $\Omega = X \times Y$, and think of α as a measure (or function) $\alpha : \Omega^k \to \mathbf{R}$. The projection α^i is now a measure (or function) $\alpha^i : X \times Y \to \mathbf{R}$.

In particular we will be interested in the projections $\bar{\pi}^i$.

For a probability measure $\gamma : X \times Y \to \mathbf{R}^+$, we will be interested, from time to time, in the γ-probability of an element $x \in X$. We denote this probability by $\gamma(x)$; thus

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

For simplicity we use the same notation $\gamma(y)$ for the γ-probability of an element $y \in Y$. The difference will be that we use the letters x, x_i, a, a_i, for elements of X, and the letters y, y_i, b, b_i, for elements of Y. It will always be clear if the element is an element of X or an element of Y.

When X, Y, U, V, Q are fixed, we will be interested in the value of the game, $w(\gamma)$, as a function of the measure $\gamma : X \times Y \to \mathbf{R}^+$. First notice that w is a continuous function. Also, if for γ_1, γ_2, the L_1 distance satisfies $\| \gamma_1 - \gamma_2 \|_1 \le \epsilon$ then $|w(\gamma_1) - w(\gamma_2)| \le \epsilon$. This is true since every protocol for one of the measures can be viewed as a protocol for the other one, with value different by at most ϵ. Thus w has a Lipschitz constant of 1.

If $\gamma = p\gamma_1 + (1-p)\gamma_2$, for $0 \le p \le 1$, then

$$w(\gamma) \le p w(\gamma_1) + (1-p) w(\gamma_2).$$

This is true because every protocol for γ can be viewed as a protocol for γ_1 and γ_2. Thus, the function w is concave. If $\gamma = \sum_{i=1}^{m} p_i\gamma_i$, where $\forall i : 0 \le p_i \le 1$, and $\sum_{i=1}^{m} p_i = 1$, then

$$w(\gamma) \le \sum_{i=1}^{m} p_i w(\gamma_i).$$

For a game G, it is sometimes convenient to consider also probabilistic protocols. In a probabilistic protocol, the "answers" of the players can depend also on a random string. Thus the "answers" are $u(x,h), v(y,h)$ (rather than $u(x), v(y)$), where h is a random string. The value of the protocol will be the probability, over the inputs (x,y), and over the random strings that $Q(x,y,u(x,h),v(y,h)) = 1$.

However, since a probabilistic protocol can be viewed as a convex combination of deterministic ones, the value of any probabilistic protocol can be achieved by a deterministic one.

Remark. In this paper the logarithm function log is always taken base 2. The natural logarithm is denoted by ln.

1.5. Organization of the paper. The paper is organized as follows: In section 2, we show how Theorem 1.2 leads to the proof of Theorem 1.1. In section 3, we review the definition, and the basic properties of the informational divergence, a basic tool of information theory. This tool is needed for the proof of Theorem 1.2. Theorem 1.2 is proved in section 4. The proof of the main lemma is deferred to section 5, and the proof of another lemma is deferred to section 6. In section 7 several generalizations of Theorem 1.1 are shortly discussed.

We remark that two "shortcuts" can be done in the paper. First, in the special case where the measure μ is a product measure (i.e., $\mu = \mu_X \times \mu_Y$), the entire argument is much simpler, and the proof follows simply from section 2 and the beginning of section 4 (plus simpler versions of several lemmas in section 3, using entropy instead of informational divergence). Also, section 6 is not really needed for the proof of the parallel repetition conjecture, but it is needed to improve the constant in the exponent to $\log(s(G))$. A simpler proof which does not use section 6 can be given, but the constant in the exponent in that proof is much worse.

2. Proof of the parallel repetition theorem.

In this section, we show how Theorem 1.2 leads to the proof of Theorem 1.1.

Theorem 1.1 claims an upper bound for the value $w(\bar{G})$. We will prove even more: we will upper bound the value $w(\bar{G}_A)$, of the game $\bar{G}_A$, for any large product set $A = A_X \times A_Y \subset \bar{X} \times \bar{Y}$.

We will first give some intuition: The proof uses a simple induction on the dimension k. The idea is to assume by Theorem 1.2 w.l.o.g. an upper bound for $w(\bar{G}_A^1)$. Given a protocol for $\bar{G_A}$, we partition A into product subsets, according to the behavior of the protocol on the first coordinate. The size of this partition is not too big, and therefore, the average size of a subset in the partition is not too small. In many of these subsets the protocol fails to satisfy $\bar{Q}$, because it fails to satisfy $\bar{Q}^1$. We can disregard these subsets. In every other subset, the predicate $\bar{Q}$ can be thought of as a $k-1$ dimensional predicate, and we can use induction to upper bound the size of the set of points, which satisfy this predicate. By this argument, we will get a recursive bound for $w(G_A^{\otimes k})$, as a function of the dimension k, and the size $\bar{\mu}(A)$.

For the game G, define $C_G(k,r)$ to be the maximum, taken over A, of the value $w(G_A^{\otimes k})$, of a game $G_A^{\otimes k} = \bar{G}_A = \bar{G}_{\bar{\pi}}$, where $A = A_X \times A_Y$ is a product set (with

$A_X \subset \bar{X}, A_Y \subset \bar{Y}$), with

$$-\log \bar{\mu}(A) \leq r$$

(i.e., $\bar{\mu}(A) \geq 2^{-r}$). Here, k is the dimension and $r \geq 0$ is real. For $k = 0$, it will be convenient to define $C_G(0, r) = 1$. We will prove an upper bound for $C_G(k, r)$ as a function of $w(G)$. The theorem will follow by taking $r = 0$.

Recall that X, Y, U, V, Q, μ refer to the game G from Theorem 1.1 and that $s = s(G)$. Assume for simplicity that $0 < w(G) < 1$ (otherwise the game is trivial). Take $0 < \Delta < 1$, (Δ will be determined later on). For $r \leq \Delta k$, take $A = A_X \times A_Y \subset \bar{X} \times \bar{Y}$, with $-\log \bar{\mu}(A) \leq r$, which achieves $C_G(k, r)$. Thus,

$$w(\bar{G}_{\bar{\pi}}) = C_G(k, r).$$

By Theorem 1.2, there exists i with

$$w(\bar{G}^i_{\bar{\pi}}) = w_i(\bar{\pi}) \leq W_2(w(G)) + c_0 \Delta^{1/16}$$

where W_2, c_0 are taken from Theorem 1.2. Without loss of generality assume that $i = 1$. Denote

$$\hat{w} = W_2(w(G)) + c_0 \Delta^{1/16}.$$

We will later on assume that Δ is such that

$$0 < \hat{w} < 1$$

and

$$2^{-\frac{1}{2}} < \hat{w}^{\frac{1}{2(\log(s)+\Delta)}} < 1$$

(at this point, the reader can ignore these two assumptions).

Let $u : \bar{X} \to \bar{U}$, $v : \bar{Y} \to \bar{V}$ be a protocol for $\bar{G}_{\bar{\pi}}$, achieving the value $w(\bar{G}_{\bar{\pi}})$. The pair $(x_1, u_1(x))$ is a function of $x \in \bar{X}$. Partition A_X according to $(x_1, u_1(x))$. Formally, $\forall x' \in X, u' \in U$ define

$$A_X(x', u') = \{x \in A_X \ \mid \ x_1 = x', u_1(x) = u'\}.$$

Then the family $\{A_X(x', u')\}$ is a partition of A_X. In the same way, define

$$A_Y(y', v') = \{y \in A_Y \ \mid \ y_1 = y', v_1(y) = v'\}.$$

Then the family $\{A_Y(y', v')\}$ is a partition of A_Y. For simplicity, denote in all the following $Z = X \times Y \times U \times V$ and $z = (x', y', u', v') \in Z$. For all $z = (x', y', u', v') \in Z$ denote

$$A(z) = A_X(x', u') \times A_Y(y', v').$$

Then the family $\{A(z)\}$ is a partition of A, and we have

$$A = \bigcup_{z \in Z} A(z).$$

For all $z \in Z$, define

$$B(z) = \left\{(x, y) \in A(z) \ \middle| \ \bar{Q}(x, y, u(x), v(y)) = 1\right\},$$

i.e., $B(z)$ is just the set of elements of $A(z)$ satisfying $\bar{Q}$. Notice that for $z \notin Q$ (i.e., z such that $Q(z) = 0$) we have $(x,y) \in A(z) \Rightarrow Q(x_1, y_1, u_1(x), v_1(y)) = Q(z) = 0$. Therefore, $z \notin Q$ implies $B(z) = \emptyset$. On the other hand, for $z \in Q$ we have that $(x,y) \in A(z) \Rightarrow Q(x_1, y_1, u_1(x), v_1(y)) = Q(z) = 1$. Therefore, for $z \in Q$ and $(x,y) \in A(z)$,

$$\bar{Q}(x, y, u(x), v(y)) = \prod_{i=2}^{k} Q(x_i, y_i, u_i(x), v_i(y)).$$

Thus in this case, $B(z)$ is a set of elements satisfying a $k-1$-dimensional predicate. This fact enables us to use induction.

For all z define

$$\alpha(z) = \bar{\pi}(A(z)), \quad \beta(z) = \frac{\bar{\pi}(B(z))}{\bar{\pi}(A(z))}.$$

Then we have the following.

Claim 2.1.

$$\sum_{z \in Q} \alpha(z) \leq \hat{w}.$$

Proof. $u_1 : \bar{X} \to U,\ v_1 : \bar{Y} \to V$ can be viewed as a protocol for $\bar{G}^1_{\bar{\pi}}$. The value of this protocol is clearly

$$\sum_{z \in Q} \bar{\pi}(A(z)) = \sum_{z \in Q} \alpha(z),$$

but this value is at most $w(\bar{G}^1_{\bar{\pi}}) \leq \hat{w}$. □

Claim 2.2. *For all $z = (x', y', u', v') \in Q$, with $\alpha(z) > 0$,*

$$\beta(z) \leq C_G(k-1, r - \log[\alpha(z)/\mu(x', y')]).$$

Proof. First notice that if $\alpha(z) > 0$ then also $\mu(x', y') > 0$, thus the logarithm is well defined.

For $k = 1$, the claim is immediate. Assume that $k > 1$. Ignoring the first coordinate, which is fixed, $A(z)$ can be viewed as a set of dimension $k-1$. Formally, define $A'(z) \subset X^{k-1} \times Y^{k-1}$ by

$$A'(z) = \left\{ (\tilde{x}, \tilde{y}) \in X^{k-1} \times Y^{k-1} \;\middle|\; ((x', \tilde{x}), (y', \tilde{y})) \in A(z) \right\}$$

where $(x', \tilde{x})$ denotes $x \in X^k$, with $x_1 = x'$, and $(x_2, \ldots, x_k) = \tilde{x}$ (and, similarly, $(y', \tilde{y})$).

Since for $(x,y) \in A(z)$: $x_1 = x'$, and $y_1 = y'$, we have by definition

$$\mu^{\otimes k}(A(z)) = \mu(x', y') \mu^{\otimes k-1}(A'(z)).$$

In the same way, define

$$B'(z) = \left\{ (\tilde{x}, \tilde{y}) \in X^{k-1} \times Y^{k-1} \;\middle|\; ((x', \tilde{x}), (y', \tilde{y})) \in B(z) \right\}.$$

Then we have

$$\mu^{\otimes k}(B(z)) = \mu(x', y') \mu^{\otimes k-1}(B'(z)).$$

The last $k-1$ coordinates of u: $\bar{X} \to \bar{U}, v$: $\bar{Y} \to \bar{V}$ can be viewed as a protocol for the game $G^{\otimes k-1}_{A'(z)}$. Since $z \in Q$, this protocol satisfies $Q^{\otimes k-1}$ at the set of elements $B'(z)$. Therefore, the value of this protocol is

$$\frac{\mu^{\otimes k-1}(B'(z))}{\mu^{\otimes k-1}(A'(z))} = \frac{\bar{\mu}(B(z))}{\bar{\mu}(A(z))} = \frac{\bar{\mu}(B(z))/\bar{\mu}(A)}{\bar{\mu}(A(z))/\bar{\mu}(A)} = \frac{\bar{\pi}(B(z))}{\bar{\pi}(A(z))} = \beta(z),$$

so by the definition of C_G we have

$$\begin{aligned}\beta(z) &\le C_G(k-1, -\log \mu^{\otimes k-1}(A'(z)) = C_G(k-1, -\log[\bar{\mu}(A(z)/\mu(x', y')]) \\ &= C_G(k-1, -\log[\bar{\mu}(A)\alpha(z)/\mu(x', y')]) \le C_G(k-1, r - \log\alpha(z) + \log\mu(x', y'))\end{aligned}$$

(recall that by definition $C_G(k', r')$ is monotone in r'). Notice that since $\mu^{\otimes k-1}(A'(z)) \le 1$, $r - \log\alpha(z) + \log\mu(x', y') \ge 0$, thus C_G is defined. □

CLAIM 2.3.

$$C_G(k, r) = \sum_{z\in Q} \alpha(z)\beta(z).$$

Proof. The protocol u, v satisfies $\bar{Q}$ at the set of elements $\bigcup_{z\in Z} B(z)$. Therefore,

$$C_G(k, r) = w(\bar{G}_{\bar{\pi}}) = \bar{\pi}\left(\bigcup_{z\in Z} B(z)\right) = \sum_{z\in Z} \bar{\pi}(B(z)) = \sum_{z\in Z} \bar{\pi}(A(z))\frac{\bar{\pi}(B(z))}{\bar{\pi}(A(z))} = \sum_{z\in Z} \alpha(z)\beta(z)$$

but $z \notin Q$ implies $\beta(z) = 0$. Thus,

$$C_G(k, r) = \sum_{z\in Q} \alpha(z)\beta(z). \qquad □$$

Denote

$$T = \{z \in Q \mid \alpha(z) > 0\}.$$

From Claims 2.2 and 2.3 we have the recursive inequality

$$C_G(k, r) \le \sum_{z\in T} \alpha(z) C_G(k-1, r - \log[\alpha(z)/\mu(x', y')]) \tag{1}$$

where, by Claim 2.1,

$$\sum_{z\in T} \alpha(z) \le \hat{w}.$$

We will now assume that Δ is such that

$$0 < \hat{w} < 1$$

and

$$2^{-1/2} < \hat{w}^{1/(2(\log(s)+\Delta))} < 1.$$

We will prove by induction on k the following inequality.

CLAIM 2.4.

$$C_G(k,r) \le \left(\hat{w}^{1/(2(\log(s)+\Delta))}\right)^{\Delta k - r}.$$

Proof. For $k=0$, the claim is trivial, (since $0<\hat{w}<1$, and $r \ge 0$). For $k \ge 1$ assume the inequality for $k-1$, and substitute in inequality (1) to get

$$\begin{aligned} C_G(k,r) &\le \sum_{z\in T} \alpha(z) \left(\hat{w}^{1/(2(\log(s)+\Delta))}\right)^{\Delta(k-1)-r+\log[\alpha(z)/\mu(x',y')]} \\ &= \left(\hat{w}^{1/(2(\log(s)+\Delta))}\right)^{\Delta k - r} \sum_{z\in T} \alpha(z) \left(\hat{w}^{1/(2(\log(s)+\Delta))}\right)^{-\Delta+\log[\alpha(z)/\mu(x',y')]}. \end{aligned}$$

Hence, it will be enough to prove

$$\sum_{z\in T} \alpha(z) \left(\hat{w}^{1/(2(\log(s)+\Delta))}\right)^{-\Delta-\log(s)+\log[s\alpha(z)/\mu(x',y')]} \ \le\ 1.$$

Define

$$t(z) = \begin{cases} 0 & \text{for } z \notin T, \\ \dfrac{s\alpha(z)}{\mu(x',y')} & \text{for } z \in T. \end{cases}$$

Then the inequality is equivalent to

$$\sum_{z\in T} \left(\frac{\mu(x',y')}{s}\right) t(z) \left(\hat{w}^{1/(2(\log(s)+\Delta))}\right)^{\log t(z)} \ \le\ \hat{w}^{1/2}.$$

Define $p(z) = \mu(x',y')/s$, and $f : \mathbf{R}^+ - \{0\} \to \mathbf{R}$ by

$$f(t) = tc^{\log t} = t^{1+\log c}$$

where

$$c = \hat{w}^{1/(2(\log(s)+\Delta))}.$$

In these notations, we have to prove

$$\sum_{z\in T} p(z) f(t(z)) \ \le\ \hat{w}^{1/2}.$$

We assumed $2^{-1/2} < c < 1$. Therefore, $-\frac{1}{2} < \log c < 0$, thus $f(t) = tc^{\log t} = t^{1+\log c}$ is convex. Notice that

$$\sum_{z\in Z} p(z) = \sum_{U\times V} \sum_{X\times Y} \frac{\mu(x',y')}{s} = \frac{1}{s} \sum_{U\times V} \sum_{X\times Y} \mu(x',y') = \frac{1}{s} \sum_{U\times V} 1 = \frac{1}{s} s = 1.$$

Therefore, we can use Jensen's inequality to conclude

$$\sum_{z\in T} p(z) f(t(z)) = \sum_{z\in T} p(z) t(z)^{1+\log c} = \sum_{z\in Z} p(z) t(z)^{1+\log c} \le \left(\sum_{z\in Z} p(z) t(z)\right)^{1+\log c},$$

but

$$\sum_{z\in Z} p(z)t(z) = \sum_{z\in T} p(z)t(z) = \sum_{z\in T} \frac{\mu(x',y')}{s}\frac{s\alpha(z)}{\mu(x',y')} = \sum_{z\in T}\alpha(z) \leq \hat{w}$$

and since the function $t^{1+\log c}$ is monotone in t, we have

$$\sum_{z\in T} p(z)f(t(z)) \leq \left(\sum_{z\in Z} p(z)t(z)\right)^{1+\log c} \leq \hat{w}^{1+\log c} \leq \hat{w}^{1-1/2} = \hat{w}^{1/2}$$

(where the third inequality uses the assumption $\log c > -1/2$). □

By Claim 2.4 and by $\Delta < 2\log_2(s)$ (which follows from the assumptions: $\Delta < 1$, and $s \geq 2$), we can now conclude

$$w(G^{\otimes k}) = C_G(k,0) \leq \left(\hat{w}^{1/(2(\log(s)+\Delta))}\right)^{\Delta k} \leq \left(\hat{w}^{1/(4\log(s))}\right)^{\Delta k} = \left(\hat{w}^{(1/4)\Delta}\right)^{k/\log(s)}$$

where $\hat{w} = W_2(w(G)) + c_0\Delta^{1/16}$, and $0 < \Delta < 1$ satisfies $0 < \hat{w} < 1$, and

$$2^{-1/2} < \hat{w}^{1/(2(\log(s)+\Delta))} < 1.$$

Since $0 < W_2(w(G)) < 1$, and since $\hat{w}$ is monotone in Δ, the conditions are satisfiable. Just start from $\Delta = 0$ and increase Δ until the conditions hold.

Thus Theorem 1.1 follows. We will take in Theorem 1.1

$$W(w(G)) = \inf\left(\hat{w}^{(1/4)\Delta}\right)$$

where the infimum is taken over all $0 < \Delta < 1$, which satisfy the conditions.

3. Informational divergence.

In this section, we define the informational divergence, a basic tool of information theory, and review some of its basic properties that will be used in the paper. The reader can find excellent treatments of the subject in [29, 17].

Given a finite probability space Ω, with two probability measures ϑ, ψ, the divergence of ϑ with respect to ψ is defined by

$$\mathbf{D}\left(\vartheta \parallel \psi\right) = \sum_{z\in\Omega} \vartheta(z)\log\frac{\vartheta(z)}{\psi(z)}.$$

In this definition, $0\log\frac{0}{0}$ is defined to be 0, and for $z > 0$, $z\log\frac{z}{0}$ is defined to be ∞. Thus $\mathbf{D}\left(\vartheta \parallel \psi\right) < \infty$ if and only if ϑ is absolutely continuous with respect to ψ (i.e., $\psi(z) = 0$ implies $\vartheta(z) = 0$).

In the special case where ψ is the uniform distribution, we have

$$\mathbf{D}\left(\vartheta \parallel \psi\right) = \log_2|\Omega| - \mathbf{H}(\vartheta),$$

where $\mathbf{H}(\vartheta)$ is the standard entropy of ϑ. More generally, the divergence $\mathbf{D}\left(\vartheta \parallel \psi\right)$ can be thought of as the entropy of the measure ϑ, relative to the measure ψ, as opposed to the standard entropy of ϑ, which is taken relative to the uniform distribution.

$\mathbf{D}\left(\vartheta \parallel \psi\right)$ has many names and notations throughout the literature. In [29] it is also called "relative entropy," and denoted by $\mathbf{H}_{\vartheta\parallel\psi}(\mathcal{Q})$, where $\mathcal{Q}$ is the partition of Ω

into single points. This is a special case of the following definition: For a measurement f on Ω, with a finite alphabet A, let $\mathcal{Q}$ be the induced partition $\{f^{-1}(a)\}_{a\in A}$. Let ϑ_f, ψ_f be the corresponding probability mass functions, i.e., for $a \in A$

$$\vartheta_f(a) = \vartheta(\{z \in \Omega \mid f(z) = a\}), \quad \psi_f(a) = \psi(\{z \in \Omega \mid f(z) = a\}).$$

The relative entropy of f, with measure ϑ, with respect to the measure ψ, is defined by

$$\mathbf{H}_{\vartheta\|\psi}(f) = \mathbf{H}_{\vartheta\|\psi}(\mathcal{Q}) = \sum_{a\in A} \vartheta_f(a) \log \frac{\vartheta_f(a)}{\psi_f(a)}.$$

In this paper, we prefer to use the notation $\mathbf{D}\left(\vartheta \parallel \psi\right)$.

The following lemma, known as the divergence inequality, is probably the most basic property of the informational divergence.

LEMMA 3.1. *For all ϑ, ψ, we have*

$$\mathbf{D}\left(\vartheta \parallel \psi\right) \geq 0.$$

Proof. See [29, Chapter 2, Theorem 2.3.1]. □

For measures $\bar{\vartheta}, \bar{\psi}$ on $\otimes_{i=1}^k \Omega_i$, recall that $\bar{\vartheta}^i, \bar{\psi}^i$ are the projections of $\bar{\vartheta}, \bar{\psi}$ on the ith coordinate.

LEMMA 3.2. *For measures $\bar{\vartheta}, \bar{\psi}$ on $\Omega_1 \times \Omega_2$, such that $\bar{\psi} = \bar{\psi}^1 \times \bar{\psi}^2$,*

$$\mathbf{D}\left(\bar{\vartheta} \parallel \bar{\psi}\right) \geq \mathbf{D}\left(\bar{\vartheta}^1 \parallel \bar{\psi}^1\right) + \mathbf{D}\left(\bar{\vartheta}^2 \parallel \bar{\psi}^2\right).$$

Proof. See [29, Chapter 2, Lemma 2.5.3]. The lemma is stated there as: $M_{XY} = M_X \times M_Y$ implies

$$\mathbf{H}_{P\|M}(X, Y) \geq \mathbf{H}_{P\|M}(X) + \mathbf{H}_{P\|M}(Y). \qquad \square$$

The next lemma can be viewed as a generalization of the well-known entropy inequality

$$\mathbf{H}(z_1, \ldots, z_k) \leq \mathbf{H}(z_1) + \cdots + \mathbf{H}(z_k)$$

(for any random variable z), and as a generalization of the previous lemma.

LEMMA 3.3. *For measures $\bar{\vartheta}, \bar{\psi}$ on Ω^k, such that $\bar{\psi} = \otimes_{i=1}^k \bar{\psi}^i$*

$$\mathbf{D}\left(\bar{\vartheta} \parallel \bar{\psi}\right) \geq \sum_{i=1}^k \mathbf{D}\left(\bar{\vartheta}^i \parallel \bar{\psi}^i\right).$$

Proof. The proof is immediate from Lemma 3.2. □

For a function $\alpha : \Omega \to \mathbf{R}$, denote by $\parallel \alpha \parallel_1$ the standard L_1 norm of α, i.e.,

$$\parallel \alpha \parallel_1 = \parallel \alpha(z) \parallel_1 = \sum_{z\in\Omega} |\alpha(z)|.$$

If α is a probability measure then $\parallel \alpha \parallel_1 = 1$. In this paper we use $\parallel \vartheta - \psi \parallel_1$ as a distance function between measures (or functions). It is not true that the divergence $\mathbf{D}\left(\vartheta \parallel \psi\right)$ is a distance function. However, it is true that if $\mathbf{D}\left(\vartheta \parallel \psi\right)$ is small then the L_1 distance between ϑ and ψ is also small.

LEMMA 3.4. *For all ϑ, ψ, we have*

$$(2\ln 2)\mathbf{D}\left(\vartheta \parallel \psi\right) \;\geq\; (\parallel \vartheta - \psi \parallel_1)^2.$$

Proof. See [17, Chapter 3, Exercise 17] and the references therein. □

The next lemma computes the value of $\mathbf{D}\left(\vartheta \parallel \psi\right)$, in the special case $\vartheta = \psi_A$, where $A \subset \Omega$.

LEMMA 3.5. *For a measure ψ, and a set $A \subset \Omega$, we have*

$$\mathbf{D}\left(\psi_A \parallel \psi\right) = -\log \psi(A).$$

Proof.

$$\mathbf{D}\left(\psi_A \parallel \psi\right) = \sum_{z\in\Omega} \psi_A(z) \log \frac{\psi_A(z)}{\psi(z)} \;=\; \sum_{z\in A} \psi_A(z) \log \frac{\psi_A(z)}{\psi(z)}$$

$$= \sum_{z\in A} \psi_A(z) \log \frac{\psi(z)/\psi(A)}{\psi(z)} \;=\; \sum_{z\in A} \psi_A(z)(-\log \psi(A)) \;=\; -\log \psi(A). \quad \square$$

For a probability measure $\alpha : \Omega \to \mathbf{R}^+$, where $\Omega = X \times Y$, define $\alpha(a,\cdot) : Y \to \mathbf{R}^+$ to be the probability measure on Y, derived from α by fixing $x = a$. Thus, $\alpha(a,\cdot)$ is the following probability measure: for all $y \in Y$,

$$\alpha(a,\cdot)(y) = \frac{\alpha(a,y)}{\alpha(a)}$$

where $\alpha(a) = \sum_{y\in Y} \alpha(a,y)$. This definition makes sense only if $\alpha(a) > 0$. Otherwise, define $\alpha(a,\cdot)$ to be identically 0.

In the same way, define the measure $\alpha(\cdot,y) : X \to \mathbf{R}^+$.

For the measures $\vartheta, \psi : X \times Y \to \mathbf{R}^+$, we will be interested in the values of $\mathbf{D}\left(\vartheta(x,\cdot) \parallel \psi(x,\cdot)\right)$, and $\mathbf{D}\left(\vartheta(\cdot,y) \parallel \psi(\cdot,y)\right)$. Define

$$\mathbf{V_X}\left(\vartheta \parallel \psi\right) = \sum_{x\in X} \vartheta(x)\mathbf{D}\left(\vartheta(x,\cdot) \parallel \psi(x,\cdot)\right),$$

and

$$\mathbf{V_Y}\left(\vartheta \parallel \psi\right) = \sum_{y\in Y} \vartheta(y)\mathbf{D}\left(\vartheta(\cdot,y) \parallel \psi(\cdot,y)\right).$$

These are denoted in [29] by $\mathbf{H}_{\vartheta\parallel\psi}(Y|X)$, and $\mathbf{H}_{\vartheta\parallel\psi}(X|Y)$. In addition, define

$$\mathbf{V}\left(\vartheta \parallel \psi\right) = \frac{1}{2}[\mathbf{V_X}\left(\vartheta \parallel \psi\right) + \mathbf{V_Y}\left(\vartheta \parallel \psi\right)].$$

The notion $\mathbf{V}\left(\vartheta \parallel \psi\right)$ is central in the rest of the paper. In particular, we will be interested in cases were $\mathbf{V}\left(\vartheta \parallel \psi\right)$ is small. We saw before that if $\mathbf{D}\left(\vartheta \parallel \psi\right)$ is small then $\parallel \vartheta - \psi \parallel_1$ is also small. Is the same true for $\mathbf{V}\left(\vartheta \parallel \psi\right)$? Taking $X = Y = \{0,1\}$ and

$$\psi = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}, \quad \vartheta = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

we have $\mathbf{V}\left(\vartheta \parallel \psi\right) = 0$, but still $\parallel \vartheta - \psi \parallel_1 = \frac{1}{2}$. Hence, in general it is not the case.

A measure $\psi : X \times Y \to \mathbf{R}^+$ is called irreducible, if there are no nontrivial partitions $X = X_1 \cup X_2$, $Y = Y_1 \cup Y_2$ such that $\psi(X_1 \times Y_2) = \psi(X_2 \times Y_1) = 0$. Every measure $\psi : X \times Y \to \mathbf{R}^+$ decomposes into its irreducible components. In general, it is true that if $\mathbf{V}(\vartheta \parallel \psi) = 0$ then ϑ has the same components as ψ and behaves like ψ on each one of them, but the ϑ-probability of each component can be different from the ψ-probability.

For irreducible ψ it can be shown that if $\mathbf{V}(\vartheta \parallel \psi) = 0$ then $\vartheta = \psi$, and that if ψ is fixed $\mathbf{V}(\vartheta \parallel \psi) \to 0$ implies $\vartheta \to \psi$. However, this convergence is not uniform. For example, we can take $X = Y = \{0, 1\}$, and

$$\psi = \begin{pmatrix} \frac{1}{2} & \epsilon \\ 0 & \frac{1}{2} - \epsilon \end{pmatrix}, \quad \vartheta = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

In this case ψ is irreducible, and $\mathbf{V}(\vartheta \parallel \psi) = O(\epsilon)$, but still $\parallel \vartheta - \psi \parallel_1 = \frac{1}{2}$.

Therefore, in this paper we will use a different characterization of measures ϑ, with small $\mathbf{V}(\vartheta \parallel \psi)$. This characterization, proved in Lemma 6.6, will intuitively say that in this case there are measures ϑ', ψ', such that ϑ' is very close to ϑ, and ψ' is very close to ψ, and such that ϑ', ψ' have the same irreducible components, and behave the same on each one of them (but not necessarily give the same mass to each component).

We remark that this characterization is not necessary to prove the parallel repetition conjecture. It is done here only to improve the constants.

4. Proof of the main theorem.

In this section we give the proof of Theorem 1.2. The proofs of two important lemmas are deferred to sections 5 and 6. Given X, Y, U, V, Q, k, define for any probability measure $\bar{\alpha} : \bar{X} \times \bar{Y} \to \mathbf{R}^+$, the following games:

1. The game $G_{\bar{\alpha}^i}$, consisting of X, Y, U, V, Q, with the measure $\bar{\alpha}^i$.
   Denote the value of this game by $w(\bar{\alpha}^i)$.
2. The game $\bar{G}_{\bar{\alpha}}$, consisting of $\bar{X}, \bar{Y}, \bar{U}, \bar{V}, \bar{Q}$, with the measure $\bar{\alpha}$.
   Denote the value of this game by $\bar{w}(\bar{\alpha})$.
3. The game $\bar{G}^i_{\bar{\alpha}}$, consisting of $\bar{X}, \bar{Y}, U, V$, with the predicate $\bar{Q}^i$, defined by

$$\forall (x, y, u, v) \in \bar{X} \times \bar{Y} \times U \times V \quad : \quad \bar{Q}^i(x, y, u, v) = Q(x_i, y_i, u, v)$$

   and with the measure $\bar{\alpha}$. Denote the value of this game by $w_i(\bar{\alpha})$.

Recall that for any probability measure $\gamma : X \times Y \to \mathbf{R}^+$, we denote by $w(\gamma)$ the value of the game G_γ , consisting of X, Y, U, V, Q, γ.

Theorem 1.2 claims that if $A = A_X \times A_Y$ is large then $w_i(\bar{\pi})$ cannot be too large for all coordinates i. If A is large it is easy to show that for many coordinates i, $\bar{\pi}^i$ is very close to μ, and, therefore, $w(\bar{\pi}^i)$ is very close to $w(\mu)$ and is not too large. Hence, it could be enough to show that for some coordinates i, $w_i(\bar{\pi})$ is upper bounded by some "well behaved" function of $w(\bar{\pi}^i)$.

For any $\bar{\alpha} : \bar{X} \times \bar{Y} \to \mathbf{R}^+$, any protocol for $G_{\bar{\alpha}^i}$ defines also a corresponding protocol, with the same value, for $\bar{G}^i_{\bar{\alpha}}$. This is true because if the distribution of (x, y) is $\bar{\alpha}$ then the distribution of (x_i, y_i) is $\bar{\alpha}^i$. Thus, given an input (x, y), for the game $\bar{G}^i_{\bar{\alpha}}$, (x_i, y_i) can be used as an input for the protocol for $G_{\bar{\alpha}^i}$. The output of this protocol, can be viewed as an output for the game $\bar{G}^i_{\bar{\alpha}}$. Therefore, we always have

$$w(\bar{\alpha}^i) \leq w_i(\bar{\alpha}).$$

The other direction is false, as the the other coordinates can give a lot of information on (x_i, y_i). There is an important special case, however, in which $w(\bar{\alpha}^i) = w_i(\bar{\alpha})$.

LEMMA 4.1. *If there exist* $\alpha_1 : X \times Y \to \mathbf{R}$, $\alpha_2 : \bar{X} \to \mathbf{R}$, $\alpha_3 : \bar{Y} \to \mathbf{R}$ *such that for all* $(x, y) \in \bar{X} \times \bar{Y}$

$$\bar{\alpha}(x, y) = \alpha_1(x_i, y_i)\alpha_2(x)\alpha_3(y),$$

then

$$w(\bar{\alpha}^i) = w_i(\bar{\alpha}).$$

Proof. We will show that in this case a protocol for $\bar{G}^i_{\bar{\alpha}}$ defines a corresponding protocol, with the same value for $G_{\bar{\alpha}^i}$. This will be true because in this special case a pair (x, y), with distribution $\bar{\alpha}$, can simply be created by the two players from a pair (a_i, b_i), with distribution $\bar{\alpha}^i$.

First notice that $\alpha_1, \alpha_2, \alpha_3$ are not unique, as we can multiply α_1 and divide α_2, by the same function $f(x_i)$, as long as for all $a \in X : f(a) \neq 0$. In the same way, we can multiply α_1 and divide α_3, by the same function $g(y_i)$, as long as for all $b \in Y : g(b) \neq 0$. Therefore, we can assume w.l.o.g. that for all a_i, b_i:

$$\sum_{\{x \in \bar{X} \mid x_i = a_i\}} \alpha_2(x), \quad \text{and} \quad \sum_{\{y \in \bar{Y} \mid y_i = b_i\}} \alpha_3(y)$$

are always 0 or 1. Also, we can assume w.l.o.g. that if one of them is 0 then $\alpha_1(a_i, b_i)$ is also 0. Therefore, in this case

$$\begin{aligned}\bar{\alpha}^i(a_i, b_i) &= \sum_{\{(x,y) \in \bar{X} \times \bar{Y} \mid (x_i, y_i) = (a_i, b_i)\}} \alpha_1(a_i, b_i)\alpha_2(x)\alpha_3(y) \\ &= \alpha_1(a_i, b_i) \left(\sum_{\{x \in \bar{X} \mid x_i = a_i\}} \alpha_2(x) \right) \left(\sum_{\{y \in \bar{Y} \mid y_i = b_i\}} \alpha_3(y) \right) \; = \; \alpha_1(a_i, b_i).\end{aligned}$$

Notice that now, given $a_i \in X$, with $\bar{\alpha}^i(a_i) \neq 0$, we have $\sum_{\{x \in \bar{X} \mid x_i = a_i\}} \alpha_2(x) = 1$. Therefore, for all $a_i \in X$, with $\bar{\alpha}^i(a_i) \neq 0$, we can define a probability distribution on $\bar{X}$ by

$$\mathrm{Pr}_{a_i}(x) = \begin{cases} 0 & \text{if } x_i \neq a_i, \\ \alpha_2(x) & \text{if } x_i = a_i. \end{cases}$$

Note that $\mathrm{Pr}_{a_i}(x)$ is exactly the α-probability for x, given that $x_i = a_i$. In the same way define for $b_i \in Y$, with $\bar{\alpha}^i(b_i) \neq 0$,

$$\mathrm{Pr}_{b_i}(y) = \begin{cases} 0 & \text{if } y_i \neq b_i, \\ \alpha_3(y) & \text{if } y_i = b_i. \end{cases}$$

Given (a_i, b_i), with $\bar{\alpha}^i(a_i, b_i) \neq 0$, choose randomly $x \in \bar{X}$ according to $\mathrm{Pr}_{a_i}(x)$ and $y \in \bar{Y}$ according to $\mathrm{Pr}_{b_i}(y)$. If (a_i, b_i) are chosen with probability $\bar{\alpha}^i(a_i, b_i)$ then (x, y) are chosen with probability

$$\bar{\alpha}^i(x_i, y_i)\mathrm{Pr}_{x_i}(x)\mathrm{Pr}_{y_i}(y) = \alpha_1(x_i, y_i)\alpha_2(x)\alpha_3(y) = \bar{\alpha}(x, y).$$

Thus, if (a_i, b_i) is a random variable, with distribution $\bar{\alpha}^i$, then (x, y) is a random variable with distribution $\bar{\alpha}$.

Players I, II can use a protocol for $\bar{G}^i_{\bar{\alpha}}$ to define a probabilistic protocol for $G_{\bar{\alpha}^i}$ in the following way: Given an input pair $(a_i, b_i) \in X \times Y$ for the game $G_{\bar{\alpha}^i}$, player I chooses randomly $x \in \bar{X}$, according to $\Pr_{a_i}(x)$, and player II chooses randomly $y \in \bar{Y}$, according to $\Pr_{b_i}(y)$. Since (a_i, b_i) is a random variable with distribution $\bar{\alpha}^i$, (x, y) is a random variable with distribution $\bar{\alpha}$ and can be used as an input for $\bar{G}^i_{\bar{\alpha}}$.

Players I, II can now use the protocol for $\bar{G}^i_{\bar{\alpha}}$ on the input (x, y), as a protocol for $G_{\bar{\alpha}^i}$ on the input (a_i, b_i). Formally, if $u : \bar{X} \to U$, $v : \bar{Y} \to V$ is the protocol for $\bar{G}^i_{\bar{\alpha}}$, and h is the random string used by the players, the protocol for $G_{\bar{\alpha}^i}$ will be $u'(a_i, h) = u(x)$, $v'(b_i, h) = v(y)$, (where x, y are created from a_i, b_i, h by the previous procedure). Since

$$\bar{Q}^i(x, y, u(x), v(y)) = Q(x_i, y_i, u(x), v(y)) = Q(a_i, b_i, u'(a_i, h), v'(a_i, h)),$$

the probability that $Q(a_i, b_i, u'(a_i, h), v'(a_i, h)) = 1$ equals exactly the probability that $\bar{Q}^i(x, y, u(x), v(y)) = 1$. Since (x, y) is a random variable with distribution $\bar{\alpha}$, this probability is the value of the original protocol for $\bar{G}^i_{\bar{\alpha}}$.

Thus we proved that a protocol for $\bar{G}^i_{\bar{\alpha}}$ defines a probabilistic protocol with the same value for $G_{\bar{\alpha}^i}$. It is well known that since a probabilistic protocol can be viewed as a convex combination of deterministic ones, there must exist a deterministic protocol with the same value. □

Since the conditions of Lemma 4.1 do not hold usually, we cannot expect them to hold for the measure $\bar{\pi}$. The idea will be to represent $\bar{\pi}$ as a convex combination of measures that satisfy the conditions of Lemma 4.1 and then to deduce bounds for the values $w_i(\bar{\pi})$.

We remark that in the simpler case, in which the original measure μ is a product measure, the measure $\bar{\pi}$ does satisfy the conditions of Lemma 4.1. Had we considered only this simpler case, the entire proof would have ended here!

Recall that in all the following $\bar{\pi} = \bar{\mu}_A$, where the set A is taken from Theorem 1.2 and satisfies,

$$-\log \bar{\mu}(A) \le \Delta k.$$

The following definition is probably the most important notion in this paper. This definition enables the representation of $\bar{\pi}$ as a convex combination of measures with nice properties. A similar idea was used before in [35, 41].

A scheme M of type $\mathcal{M}^l$ consists of:

1. A partition of the set of coordinates $[k] - \{l\}$ into $I \cup J$.
2. Values $a_i \in X$, for all $i \in I$, and $b_j \in Y$, for all $j \in J$.

(Formally, M should be denoted by $M^l_{I,J,a_{i_1},\ldots,a_{i_{|I|}},b_{j_1},\ldots,b_{j_{|J|}}}$; however, for simplicity we will just use the notation M.) We also denote by M the set

$$M = \left\{ (x, y) \in \bar{X} \times \bar{Y} \;\middle|\; \forall i \in I \ : \ x_i = a_i \ , \ \ \forall j \in J \ : \ y_j = b_j \right\}.$$

We also denote by $\mathcal{M}^l$ the family of all sets M of type $\mathcal{M}^l$ (i.e., for all possible $I, J, a_{i_1}, \ldots, a_{i_{|I|}}, b_{j_1}, \ldots, b_{j_{|J|}}$). Notice that each $\mathcal{M}^l$ is a cover of $\bar{X} \times \bar{Y}$. Each element $(x, y) \in \bar{X} \times \bar{Y}$ is covered 2^{k-1} times by each $\mathcal{M}^l$.

For all l, define the following two probability measures, $\nu_l, \rho_l : \mathcal{M}^l \to \mathbf{R}^+$ by

$$\nu_l(M) = \frac{\bar{\mu}(M)}{2^{k-1}}, \quad \rho_l(M) = \frac{\bar{\pi}(M)}{2^{k-1}} = \frac{\bar{\mu}_A(M)}{2^{k-1}} = \frac{\bar{\mu}(A \cap M)}{\bar{\mu}(A) 2^{k-1}}$$

(ν_l, ρ_l are obviously probability measures, by the above observation that each element (x, y) is covered exactly 2^{k-1} times by $\mathcal{M}^l$).

Recall that for the set M, we have the measures $\bar{\mu}_M : \bar{X} \times \bar{Y} \to \mathbf{R}^+$, $\bar{\pi}_M : \bar{X} \times \bar{Y} \to \mathbf{R}^+$. For M with $\bar{\mu}(M) > 0$ (respectively, $\bar{\pi}(M) > 0$), these measures are probability measures (recall that otherwise they are defined to be identically 0). Recall that $\bar{\mu}^i_M, \bar{\pi}^i_M : X \times Y \to \mathbf{R}^+$ are the projections of these measures on the ith coordinate. For $M \in \mathcal{M}^l$, we will be mainly interested in the projections $\bar{\mu}^l_M, \bar{\pi}^l_M$. Note that since $\bar{\mu} = \mu^{\otimes k}$, the projection $\bar{\mu}^l_M$ is just μ.

The important fact for M is the following.

CLAIM 4.1. *For $M \in \mathcal{M}^l$ (with $\rho_l(M) > 0$),*

$$w_l(\bar{\pi}_M) = w(\bar{\pi}^l_M).$$

Proof. This is a simple application of Lemma 4.1. For $(x, y) \in M$, we have

$$\begin{aligned}\bar{\mu}_M(x, y) = \bar{\mu}(M)^{-1}\bar{\mu}(x, y) \ &= \ \bar{\mu}(M)^{-1}\prod_{i=1}^{k}\mu(x_i, y_i)\\ &= \bar{\mu}(M)^{-1}\mu(x_l, y_l)\prod_{i\in I}\mu(x_i, y_i)\prod_{j\in J}\mu(x_j, y_j)\\ &= \bar{\mu}(M)^{-1}\mu(x_l, y_l)\prod_{i\in I}\mu(a_i, y_i)\prod_{j\in J}\mu(x_j, b_j).\end{aligned}$$

Define $\mu_1 : X \times Y \to \mathbf{R}$, $\mu_2 : \bar{X} \to \mathbf{R}$, $\mu_3 : \bar{Y} \to \mathbf{R}$ by

$$\mu_1(x_l, y_l) = \bar{\mu}(M)^{-1}\mu(x_l, y_l), \ \ \mu_2(x) = \prod_{j\in J}\mu(x_j, b_j), \ \ \mu_3(y) = \prod_{i\in I}\mu(a_i, y_i).$$

Then for $(x, y) \in M$, we have

$$\bar{\mu}_M(x, y) = \mu_1(x_l, y_l)\mu_2(x)\mu_3(y).$$

The event $(x, y) \in M \cap A$ can be written as

$$(x \in A_X \ \cap \ \forall i \in I : x_i = a_i) \bigcap (y \in A_Y \ \cap \ \forall j \in J : y_j = b_j).$$

Therefore, defining $\chi_2 : \bar{X} \to \mathbf{R}$, $\chi_3 : \bar{Y} \to \mathbf{R}$ by

$$\chi_2(x) = \begin{cases} 1 & \text{if } x \in A_X \ \cap \ \forall i \in I \ : \ x_i = a_i, \\ 0 & \text{otherwise,} \end{cases}$$

$$\chi_3(y) = \begin{cases} 1 & \text{if } y \in A_Y \ \cap \ \forall j \in J \ : \ y_j = b_j, \\ 0 & \text{otherwise,} \end{cases}$$

we have $\forall (x, y) \in \bar{X} \times \bar{Y}$

$$\bar{\pi}_M(x, y) = \bar{\mu}_{A\cap M}(x, y) = (\bar{\mu}_M)_A(x, y) = \bar{\mu}_M(A)^{-1}\chi_2(x)\chi_3(y)\bar{\mu}_M(x, y),$$

and since $(x, y) \notin M$ implies $\chi_2(x)\chi_3(y) = 0$, we have

$$\bar{\pi}_M(x, y) = \bar{\mu}_M(A)^{-1}\chi_2(x)\chi_3(y)\mu_1(x_l, y_l)\mu_2(x)\mu_3(y).$$

Define

$$\begin{aligned}\pi_1(x_l, y_l) &= \bar{\mu}_M(A)^{-1}\mu_1(x_l, y_l),\\ \pi_2(x) &= \chi_2(x)\mu_2(x),\\ \pi_3(y) &= \chi_3(y)\mu_3(y).\end{aligned}$$

Then $\bar{\pi}_M(x, y) = \pi_1(x_l, y_l)\pi_2(x)\pi_3(y)$ and the claim follows from Lemma 4.1. □

Recall the definitions of the probability measures ν_l, ρ_l. Since every (x, y) is covered the same number of times by $\mathcal{M}^l$, for any l, the measure $\bar{\pi}$ can be written as the convex combination

$$\bar{\pi} = \sum_{M \in \mathcal{M}^l} \rho_l(M) \bar{\pi}_M. \tag{2}$$

We will use the equality (2) in two ways. First note that a protocol for $\bar{G}^l_{\bar{\pi}}$ is also a protocol for each $\bar{G}^l_{\bar{\pi}_M}$. Therefore,

$$w_l(\bar{\pi}) \leq \sum_{M \in \mathcal{M}^l} \rho_l(M) w_l(\bar{\pi}_M) = \mathbf{E}_{\rho_l}\left(w_l(\bar{\pi}_M)\right),$$

(which reflects the concavity of the function w_l–see in the Introduction), and by Claim 4.1 we have

$$w_l(\bar{\pi}) \leq \mathbf{E}_{\rho_l}\left(w\left(\bar{\pi}^l_M\right)\right). \tag{3}$$

Thus, if we can only upper bound the values $w\left(\bar{\pi}^l_M\right)$, of the one-dimensional measures $\bar{\pi}^l_M$, we will have an upper bound for $w_l(\bar{\pi})$.

In order to deduce an upper bound for $\mathbf{E}_{\rho_l}\left(w\left(\bar{\pi}^l_M\right)\right)$, we need some more properties of the family $\left\{\bar{\pi}^l_M\right\}_{M \in \mathcal{M}^l}$: First take the projection of equality (2) to get

$$\bar{\pi}^l = \sum_{M \in \mathcal{M}^l} \rho_l(M) \bar{\pi}^l_M.$$

We also need the following lemma.

Lemma 4.2 (main). *There exists l_0 such that*

$$\mathbf{E}_{\rho_{l_0}}\left(\mathbf{V}\left(\bar{\pi}^{l_0}_M \,\middle\|\, \mu\right)\right) \leq \frac{2}{k}(-\log \bar{\mu}(A))$$

and

$$\mathbf{D}\left(\bar{\pi}^{l_0} \,\|\, \mu\right) \leq \frac{2}{k}(-\log \bar{\mu}(A)).$$

The proof is given in the next section.

In order to understand the intuition behind Lemma 4.2, one should think of the right-hand side, $\frac{2}{k}(-\log \bar{\mu}(A))$, as a very small ϵ. The lemma just says that there exists l_0 s.t. $\bar{\pi}^{l_0}$ is very close to μ (by Lemma 3.4), and s.t. for an average M, $\mathbf{V}(\bar{\pi}^{l_0}_M \| \mu)$ is very small.

Thus fixing l_0 from the last lemma, and fixing $\epsilon = 2/k(-\log \bar{\mu}(A))$, the family of measures $\{\bar{\pi}^{l_0}_M\}_{M \in \mathcal{M}^{l_0}}$ satisfies

$$\sum_{M \in \mathcal{M}^{l_0}} \rho_{l_0}(M) \mathbf{V}\left(\bar{\pi}^{l_0}_M \,\middle\|\, \mu\right) \;\leq\; \epsilon$$

and

$$\mathbf{D}\left(\sum_{M \in \mathcal{M}^{l_0}} \rho_{l_0}(M) \bar{\pi}^{l_0}_M \,\middle\|\, \mu\right) = \mathbf{D}\left(\bar{\pi}^{l_0} \,\|\, \mu\right) \;\leq\; \epsilon.$$

In these conditions, the following lemma proves an upper bound for

$$\sum_{M \in \mathcal{M}^{l_0}} \rho_{l_0}(M) w\left(\bar{\pi}_M^{l_0}\right) = \mathbf{E}_{\rho_{l_0}}\left(w\left(\bar{\pi}_M^{l_0}\right)\right),$$

i.e., an upper bound for the value $w\left(\bar{\pi}_M^{l_0}\right)$, for an average M.

LEMMA 4.3. *There exists a global function $W_2 : [0,1] \to [0,1]$, with $z < 1$ implying $W_2(z) < 1$, and a global constant c_1 such that: if $\{\gamma_d\}_{d \in D}$ is a family of probability measures on $X \times Y$, and $\{p_d\}_{d \in D}$ are weights satisfying $\forall d \,:\, p_d \geq 0$, and $\sum_{d \in D} p_d = 1$. Assume that for some $0 \leq \epsilon \leq 1$*

$$\sum_D p_d \mathbf{V}\left(\gamma_d \parallel \mu\right) \;\leq\; \epsilon$$

and

$$\mathbf{D}\left(\sum_D p_d \gamma_d \,\middle\|\, \mu\right) \;\leq\; \epsilon.$$

Then,

$$\sum_D p_d w(\gamma_d) \;\leq\; W_2(w(\mu)) + c_1 \epsilon^{1/16}.$$

The proof is given in section 6.

Using Lemma 4.3, and inequality (3), and the fact that

$$\epsilon \overset{\text{def}}{=} \frac{2}{k}(-\log \bar{\mu}(A)) \;\leq\; \frac{2}{k}\Delta k \;\leq\; 2\Delta,$$

we can now conclude

$$w_{l_0}(\bar{\pi}) \;\leq\; \mathbf{E}_{\rho_{l_0}}\left(w\left(\bar{\pi}_M^{l_0}\right)\right) \;\leq\; W_2(w(\mu)) + c_1(2\Delta)^{1/16},$$

which proves Theorem 1.2.

Lemma 4.2, which is the most important lemma in the paper, is proved in the next section. Lemma 4.3 is proved in section 6.

We remark that Lemma 4.3 is not really necessary: for small enough ϵ, and irreducible μ, the fact that $\mathbf{V}\left(\gamma_d \parallel \mu\right) \leq \epsilon$ implies that $\parallel \gamma_d - \mu \parallel_1$ is very small. Therefore, it can be proved easily that

$$\mathbf{E}_{\rho_{l_0}}\left(w\left(\bar{\pi}_M^{l_0}\right)\right) \approx w(\mu),$$

which implies the parallel repetitions conjecture for irreducible measures. It turns out that the general case follows easily from the irreducible one.

However, as we mentioned in section 3, only for a very small ϵ it is the case that $\mathbf{V}\left(\gamma_d \parallel \mu\right) \leq \epsilon$ implies that $\parallel \gamma_d - \mu \parallel_1$ is small. In fact, such an ϵ needs to be much smaller than the ϵ used here. In particular, since the measure μ is arbitrary, such an ϵ depends on the size of the input set, $|X \times Y|$, as opposed to the ϵ used here, which depends only on $w(\mu)$ and is independent of other parameters of the game. Thus using Lemma 4.3 improves the constants in the exponent (in our analysis) in the parallel repetition theorem.

5. The main lemma. In this section we give the proof of Lemma 4.2. Our proof uses tools, and intuition from [41].

The main idea of the proof is to introduce a new type of scheme. A scheme M of type $\mathcal{M}$ consists of

1. a partition of the set of coordinates $[k]$ into $I \cup J$;
2. values $a_i \in X$, for all $i \in I$, and $b_j \in Y$, for all $j \in J$;

Note that this is the same as before, except that here $I \cup J = [k]$ rather than $[k] - \{l\}$. As before, we also denote by M the set

$$M = \left\{ (x, y) \in \bar{X} \times \bar{Y} \;\middle|\; \forall i \in I \;:\; x_i = a_i \;,\;\; \forall j \in J \;:\; y_j = b_j \right\}$$

and by $\mathcal{M}$ the family of all the sets M of type $\mathcal{M}$. Notice that each element $(x, y) \in \bar{X} \times \bar{Y}$ is covered 2^k times, by the cover $\mathcal{M}$. As before, define two probability measures $\nu, \rho : \mathcal{M} \to \mathbf{R}^+$ by

$$\nu(M) = \frac{\bar{\mu}(M)}{2^k} \quad , \quad \rho(M) = \frac{\bar{\pi}(M)}{2^k} = \frac{\bar{\mu}_A(M)}{2^k} = \frac{\bar{\mu}(A \cap M)}{\bar{\mu}(A) 2^k}.$$

As before, we have for the set M the measures $\bar{\mu}_M : \bar{X} \times \bar{Y} \to \mathbf{R}^+$, and $\bar{\pi}_M = \bar{\mu}_{M \cap A} : \bar{X} \times \bar{Y} \to \mathbf{R}^+$. Recall that $\bar{\mu}^i_M, \bar{\pi}^i_M$ are the projections of these measures on the ith coordinate.

For $i \in I$, $x_i = a_i$ is fixed. In this case $\bar{\mu}^i_M, \bar{\pi}^i_M$ are concentrated on $\{a_i\} \times Y$ and can also be thought of as measures on Y. In the same way, if $j \in J$ then $\bar{\mu}^j_M, \bar{\pi}^j_M$ can be thought of as measures on X.

Notice also that since $\bar{\mu} = \mu^{\otimes k}$, we have

$$\bar{\mu}_M = \otimes_{i=1}^k \bar{\mu}^i_M.$$

Claim 5.1.

$$\sum_{i=1}^k \mathbf{E}_\rho \left(\mathbf{D} \left(\bar{\pi}^i_M \;\middle\|\; \bar{\mu}^i_M \right) \right) \leq -\log \bar{\mu}(A).$$

Proof. The proof follows from the basic properties of informational divergence: For any scheme $M \in \mathcal{M}$ with $\rho(M) > 0$, we have by Lemma 3.5,

$$\begin{aligned} \mathbf{D}\left(\bar{\pi}_M \parallel \bar{\mu}_M \right) &= \mathbf{D}\left(\bar{\mu}_{M \cap A} \parallel \bar{\mu}_M \right) = \mathbf{D}\left((\bar{\mu}_M)_A \parallel \bar{\mu}_M \right) \\ &= -\log \bar{\mu}_M(A) = -\log \frac{\bar{\mu}(M \cap A)}{\bar{\mu}(M)} = -\log \frac{\rho(M)\bar{\mu}(A)}{\nu(M)}. \end{aligned}$$

Therefore,

$$\begin{aligned} \mathbf{E}_\rho \left(\mathbf{D}\left(\bar{\pi}_M \parallel \bar{\mu}_M \right) \right) &= -\sum_{M \in \mathcal{M}} \rho(M) \log \frac{\rho(M)\bar{\mu}(A)}{\nu(M)} \\ &= -\sum_{M \in \mathcal{M}} \rho(M) \left[\log \frac{\rho(M)}{\nu(M)} + \log \bar{\mu}(A) \right] \\ &= -\mathbf{D}\left(\rho \parallel \nu \right) - \log \bar{\mu}(A) \;\; \leq \; -\log \bar{\mu}(A) \end{aligned}$$

(by Lemma 3.1). But by Lemma 3.3, for all M, with $\rho(M) > 0$,

$$\mathbf{D}\left(\bar{\pi}_M \parallel \bar{\mu}_M \right) \geq \sum_{i=1}^k \mathbf{D}\left(\bar{\pi}^i_M \;\middle\|\; \bar{\mu}^i_M \right)$$

and we can conclude that

$$\sum_{i=1}^{k} \mathbf{E}_\rho \left(\mathbf{D}\left(\bar{\pi}_M^i \;\|\; \bar{\mu}_M^i\right)\right) = \mathbf{E}_\rho \left(\sum_{i=1}^{k} \mathbf{D}\left(\bar{\pi}_M^i \;\|\; \bar{\mu}_M^i\right)\right)$$
$$\leq \mathbf{E}_\rho \left(\mathbf{D}\left(\bar{\pi}_M \;\|\; \bar{\mu}_M\right)\right) \leq -\log \bar{\mu}(A). \qquad \square$$

Fix l. In what follows we denote by M a scheme of type $\mathcal{M}$, and by M' a scheme of type $\mathcal{M}^l$. Denote by I, J the partition of coordinates, corresponding to the scheme M, and by I', J' the one corresponding to M'. A scheme M agrees with a scheme M' if and only if $I' = I \cap ([k] - \{l\})$, $J' = J \cap ([k] - \{l\})$, and M, M' agree on the values of a_i, b_j, for all $i \in I', j \in J'$. For a scheme M' of type $\mathcal{M}^l$, denote the set of all schemes M of type $\mathcal{M}$, agreeing with M', by $N(M')$. Then the family of sets $\{M\}_{M \in N(M')}$ is a cover of the set M', where each element $(x, y) \in M'$ is covered exactly twice (as we have to choose $l \in I$ or $l \in J$ and then fix the value of a_l or b_l). Therefore, we have

$$\bar{\pi}(M') = \frac{1}{2} \sum_{M \in N(M')} \bar{\pi}(M),$$

and hence,

$$\rho_l(M') = \sum_{M \in N(M')} \rho(M).$$

Assume in all of the following that $\rho_l(M') > 0$. Then $\rho(M)/\rho_l(M')$ is a probability measure on $N(M')$. A scheme M can be randomly chosen, according to the distribution ρ, by first choosing M' according to ρ_l and then choosing $M \in N(M')$ with probability $\rho(M)/\rho_l(M')$.

Claim 5.2. *For every l,*

$$\mathbf{V}\left(\bar{\pi}_{M'}^l \;\|\; \bar{\mu}_{M'}^l\right) = \sum_{M \in N(M')} \frac{\rho(M)}{\rho_l(M')} \mathbf{D}\left(\bar{\pi}_M^l \;\|\; \bar{\mu}_M^l\right).$$

Proof. The proof follows by looking carefully into the definitions. For $M \in N(M')$

$$\frac{\rho(M)}{\rho_l(M')} = \frac{\bar{\pi}(M)/2^k}{\bar{\pi}(M')/2^{k-1}} = \frac{1}{2} \frac{\bar{\pi}(M)}{\bar{\pi}(M')} = \frac{1}{2} \bar{\pi}_{M'}(M).$$

If for the scheme M, $l \in I$, then $\bar{\pi}_{M'}(M)$ is just the probability to have $x_l = a_l$ in the set M'. By definition this probability is $\bar{\pi}_{M'}^l(a_l)$. Therefore, in this case,

$$\frac{\rho(M)}{\rho_l(M')} = \frac{1}{2} \bar{\pi}_{M'}^l(a_l).$$

Also in this case, $\bar{\pi}_M$ is derived from $\bar{\pi}_{M'}$ by fixing x_l to be a_l. Hence, $\bar{\pi}_M^l = \bar{\pi}_{M'}^l(a_l, \cdot)$. In the same way, we have $\bar{\mu}_M^l = \bar{\mu}_{M'}^l(a_l, \cdot)$. Therefore, for $l \in I$

$$\frac{\rho(M)}{\rho_l(M')} \mathbf{D}\left(\bar{\pi}_M^l \;\|\; \bar{\mu}_M^l\right) = \frac{1}{2} \bar{\pi}_{M'}^l(a_l) \mathbf{D}\left(\bar{\pi}_{M'}^l(a_l, \cdot) \;\|\; \bar{\mu}_{M'}^l(a_l, \cdot)\right).$$

In the same way, for $l \in J$

$$\frac{\rho(M)}{\rho_l(M')} \mathbf{D}\left(\bar{\pi}_M^l \;\|\; \bar{\mu}_M^l\right) = \frac{1}{2} \bar{\pi}_{M'}^l(b_l) \mathbf{D}\left(\bar{\pi}_{M'}^l(\cdot, b_l) \;\|\; \bar{\mu}_{M'}^l(\cdot, b_l)\right).$$

Partitioning $M \in N(M')$ into schemes with $l \in I$, and schemes with $l \in J$, we have

$$\begin{aligned}
&\sum_{M \in N(M')} \frac{\rho(M)}{\rho_l(M')} \mathbf{D}\left(\bar{\pi}^l_M \;\middle\|\; \bar{\mu}^l_M\right) \\
&= \sum_{a \in X} \frac{1}{2} \bar{\pi}^l_{M'}(a) \mathbf{D}\left(\bar{\pi}^l_{M'}(a,\cdot) \;\middle\|\; \bar{\mu}^l_{M'}(a,\cdot)\right) + \sum_{b \in Y} \frac{1}{2} \bar{\pi}^l_{M'}(b) \mathbf{D}\left(\bar{\pi}^l_{M'}(\cdot,b) \;\middle\|\; \bar{\mu}^l_{M'}(\cdot,b)\right) \\
&= \frac{1}{2} \mathbf{V}_{\mathbf{X}}\left(\bar{\pi}^l_{M'} \;\middle\|\; \bar{\mu}^l_{M'}\right) + \frac{1}{2} \mathbf{V}_{\mathbf{Y}}\left(\bar{\pi}^l_{M'} \;\middle\|\; \bar{\mu}^l_{M'}\right) = \mathbf{V}\left(\bar{\pi}^l_{M'} \;\middle\|\; \bar{\mu}^l_{M'}\right). \qquad \square
\end{aligned}$$

CLAIM 5.3.

$$\sum_{l=1}^{k} \mathbf{E}_{\rho_l}\left(\mathbf{V}\left(\bar{\pi}^l_{M'} \;\middle\|\; \bar{\mu}^l_{M'}\right)\right) \leq -\log \bar{\mu}(A).$$

Proof. By Claim 5.2 we have

$$\begin{aligned}
\mathbf{E}_{\rho}\left(\mathbf{D}\left(\bar{\pi}^l_M \;\middle\|\; \bar{\mu}^l_M\right)\right) &= \sum_{M \in \mathcal{M}} \rho(M) \mathbf{D}\left(\bar{\pi}^l_M \;\middle\|\; \bar{\mu}^l_M\right) \\
&= \sum_{M' \in \mathcal{M}^l} \rho_l(M') \sum_{M \in N(M')} \frac{\rho(M)}{\rho_l(M')} \mathbf{D}\left(\bar{\pi}^l_M \;\middle\|\; \bar{\mu}^l_M\right) \\
&= \sum_{M' \in \mathcal{M}^l} \rho_l(M') \mathbf{V}\left(\bar{\pi}^l_{M'} \;\middle\|\; \bar{\mu}^l_{M'}\right) = \mathbf{E}_{\rho_l}\left(\mathbf{V}\left(\bar{\pi}^l_{M'} \;\middle\|\; \bar{\mu}^l_{M'}\right)\right).
\end{aligned}$$

Since this is true for all l, we can sum up and conclude by Claim 5.1 that

$$\sum_{l=1}^{k} \mathbf{E}_{\rho_l}\left(\mathbf{V}\left(\bar{\pi}^l_{M'} \;\middle\|\; \bar{\mu}^l_{M'}\right)\right) = \sum_{l=1}^{k} \mathbf{E}_{\rho}\left(\mathbf{D}\left(\bar{\pi}^l_M \;\middle\|\; \bar{\mu}^l_M\right)\right) \leq -\log \bar{\mu}(A). \qquad \square$$

Since $\mathbf{V}(\vartheta \parallel \psi)$ is always non-negative (see section 3), we can conclude from the last claim that less than $\frac{k}{2}$ coordinates l satisfy

$$\mathbf{E}_{\rho_l}\left(\mathbf{V}\left(\bar{\pi}^l_{M'} \;\middle\|\; \bar{\mu}^l_{M'}\right)\right) > -\frac{2}{k} \log \bar{\mu}(A).$$

Therefore, more than $\frac{1}{2}k$ coordinates l satisfy

$$\mathbf{E}_{\rho_l}\left(\mathbf{V}\left(\bar{\pi}^l_{M'} \;\middle\|\; \bar{\mu}^l_{M'}\right)\right) \leq -\frac{2}{k} \log \bar{\mu}(A).$$

By Lemmas 3.3 and 3.5, we also have

$$\sum_{l=1}^{k} \mathbf{D}\left(\bar{\pi}^l \;\middle\|\; \bar{\mu}^l\right) \leq \mathbf{D}\left(\bar{\pi} \parallel \bar{\mu}\right) = \mathbf{D}\left(\bar{\mu}_A \parallel \bar{\mu}\right) = -\log \bar{\mu}(A)$$

and as before, more than $\frac{1}{2}k$ coordinates l satisfy

$$\mathbf{D}\left(\bar{\pi}^l \;\middle\|\; \bar{\mu}^l\right) \leq -\frac{2}{k} \log \bar{\mu}(A).$$

Therefore, there exists l with

$$\mathbf{E}_{\rho_l}\left(\mathbf{V}\left(\bar{\pi}^l_{M'} \,\|\, \bar{\mu}^l_{M'}\right)\right) \leq -\frac{2}{k}\log\bar{\mu}(A)$$

and

$$\mathbf{D}\left(\bar{\pi}^l \,\|\, \bar{\mu}^l\right) \leq -\frac{2}{k}\log\bar{\mu}(A).$$

Since $\bar{\mu} = \mu^{\otimes k}$, we have $\bar{\mu}^l_{M'} = \mu$ and $\bar{\mu}^l = \mu$. Thus Lemma 4.2 follows.

6. Families of local protocols.

Given X, Y, U, V, Q, every probability measure, $\gamma : X \times Y \to \mathbf{R}^+$, defines a game G_γ. In this section, as before, we denote the value of this game by $w(\gamma)$. The original measure μ is one measure of this type (with value $w(\mu)$). For some finite set D, let $\{\gamma_d\}_{d\in D}$ be a family of measures of this type. Thus, for all $d \in D$, $\gamma_d : X \times Y \to \mathbf{R}^+$ is a probability measure, and has a value $w(\gamma_d)$. Let $p_d \geq 0$ be an arbitrary weight ($\forall d \in D$). We will not always require $\sum_{d\in D} p_d = 1$. However, in the cases that we consider $\sum_{d\in D} p_d$ will be very close to 1. Thus p_d will be "almost" a probability measure on D.

Define

$$w_d = w(\gamma_d), \quad w = \sum_{d\in D} p_d w_d$$

(note that w may be larger than 1, because we didn't require $\sum_{d\in D} p_d = 1$), and the measure $\gamma : X \times Y \to \mathbf{R}^+$ by

$$\gamma = \sum_{d\in D} p_d \gamma_d.$$

In this section we assume that γ is close to μ. We would like to deduce, under some conditions on the measures $\{\gamma_d\}_{d\in D}$, a lower bound for $w(\mu)$ as a function of w, or conversely, an upper bound for w as a function of $w(\mu)$. The conditions on $\{\gamma_d\}_{d\in D}$ will intuitively say that each measure γ_d "locally" describes the behavior of the original measure μ. The goal of the section is to prove Lemma 4.3, which is stated in section 4.

6.1. The basic lemma.

First assume that for all $d \in D$, we have sets $X_d \subset X$ and $Y_d \subset Y$, such that

$$\gamma_d = \mu_{X_d \times Y_d}.$$

For all $d \in D$, define

$$a_d = \mu(X_d \times Y_d)$$

and

$$r_d = \frac{\mu\left[(X_d \times (Y - Y_d)) \cup ((X - X_d) \times Y_d)\right]}{\mu(X_d \times Y_d)}.$$

Define

$$r = \sum_{d\in D} p_d r_d.$$

The next lemma gives our basic lower bound for $w(\mu)$ as a function of w.

LEMMA 6.1. *Assume that for some constants* $0 \leq \epsilon_0 \leq 1$, $0 \leq \epsilon_1 \leq 1$, *we have* $\| \gamma - \mu \|_1 \leq \epsilon_0$, *and* $r \leq \epsilon_1$, *and assume (for simplicity) that* $w \leq 1$. *Let* $f : \mathbf{R} \rightarrow \mathbf{R}$ *be any increasing monotone function such that* $f(z) \leq 0$, *for* $z \leq 0$, *and such that for all* $0 \leq z \leq 1$, $0 < \delta \leq 1$, *we have the inequality*

$$(1-\delta)z + \delta f\left[\frac{1}{\delta}(z-(1-\delta))\right] \geq f(z).$$

Then

$$w(\mu) \geq f(w - 2\sqrt{\epsilon_1} - \epsilon_0).$$

We remark that ϵ_0, ϵ_1 should be thought of as small constants. Some examples for functions f, as above, are given in Corollaries 6.2, 6.3, 6.4. It is always true that for $0 \leq z \leq 1$, $f(z) \leq z$, (to see this, substitute $\delta = 1/2$ to get $f(z) \leq z/2 + f(2z-1)/2$, and by the fact that f is monotone, and $z \leq 1$, we get $f(z) \leq z/2 + f(z)/2$).

Proof. We will first give some intuition. Every measure μ decomposes into its irreducible components. In the simplest case, where $\epsilon_0 = \epsilon_1 = 0$, for all $d : r_d = 0$. Therefore, in this case, $X_d \times Y_d$ is a component of the measure μ (not necessarily irreducible). Since we can further decompose each one of these components, we can assume w.l.o.g. that $X_d \times Y_d$ is irreducible, and therefore that the family $\{\gamma_d\}_{d \in D}$ describes the decomposition of the measure μ into irreducible components. Every protocol for μ defines a protocol for each one of its components. Also, given a protocol for each one of the irreducible components, we can define a protocol for μ (we define the protocol for μ on each irreducible component separately). The value of the protocol for μ is the weighted average of the values of the protocols for the components. Therefore in the special case $\epsilon_0 = \epsilon_1 = 0$, we can simply conclude that $w(\mu) = w$, and the claim follows. In the general case the intuition is basically the same: We will find d_0 such that $X_{d_0} \times Y_{d_0}$ is "almost" a component of μ (i.e., d_0 with small r_{d_0}) and such that w_{d_0} is very close to w. We will concentrate on the set $(X - X_{d_0}) \times (Y - Y_{d_0})$ and continue by induction. Intuitively, we will get a subfamily of $\{\gamma_d\}_{d \in D}$ that will "almost" describe a decomposition of the measure μ into "almost" irreducible components.

First note that

$$\| \gamma \|_1 = \left\| \sum_{d \in D} p_d \gamma_d \right\|_1 = \sum_{d \in D} p_d \| \gamma_d \|_1 = \sum_{d \in D} p_d.$$

Therefore, since $\| \gamma - \mu \|_1 \leq \epsilon_0$ and since $\| \mu \|_1 = 1$, we have by the triangle inequality

$$1 - \epsilon_0 \leq \sum_{d \in D} p_d \leq 1 + \epsilon_0.$$

CLAIM 6.1. *There exists* $d_0 \in D$ *with*

$$r_{d_0} \leq \sqrt{\epsilon_1}$$

and

$$w_{d_0} \geq w - \sqrt{\epsilon_1} - \epsilon_0.$$

Proof. Denote

$$Z = \{d \in D \mid r_d \geq \sqrt{\epsilon_1}\}.$$

Then,

$$\epsilon_1 \geq r = \sum_{d \in D} p_d r_d \geq \sum_{d \in Z} p_d r_d \geq \sqrt{\epsilon_1} \sum_{d \in Z} p_d.$$

Hence,

$$\sum_{d \in Z} p_d \leq \sqrt{\epsilon_1};$$

thus,

$$w = \sum_{Z} p_d w_d + \sum_{D-Z} p_d w_d \leq \sum_{Z} p_d 1 + \sum_{D-Z} p_d w_d \leq \sqrt{\epsilon_1} + \sum_{D-Z} p_d w_d$$

and therefore,

$$\sum_{D-Z} p_d w_d \geq w - \sqrt{\epsilon_1}.$$

Let $d_0 \in D - Z$ be the element with the maximal w_{d_0}. Then,

$$w_{d_0} \sum_{D-Z} p_d \geq \sum_{D-Z} p_d w_d \geq w - \sqrt{\epsilon_1},$$

and by $\sum_{D-Z} p_d \leq \sum_{d \in D} p_d \leq 1 + \epsilon_0$, we have

$$w_{d_0} \geq (w - \sqrt{\epsilon_1})/(1+\epsilon_0) \geq (w - \sqrt{\epsilon_1})(1 - \epsilon_0) \geq w - \sqrt{\epsilon_1} - \epsilon_0$$

(since $w \leq 1$).

This completes the proof of Claim 6.1. □

Fix d_0 from the last claim. Define

$$X' = X - X_{d_0}, \quad Y' = Y - Y_{d_0}, \quad \delta = \mu(X' \times Y').$$

Assume that $\delta > 0$. For all $d \in D$ define

$$X'_d = X_d \cap X', \quad Y'_d = Y_d \cap Y', \quad \gamma'_d = \mu_{X'_d \times Y'_d}, \quad w'_d = w(\gamma'_d), \quad a'_d = \mu(X'_d \times Y'_d),$$
$$r'_d = \frac{\mu\left[(X'_d \times (Y' - Y'_d)) \cup ((X' - X'_d) \times Y'_d)\right]}{\mu(X'_d \times Y'_d)}, \quad p'_d = p_d \frac{1}{\delta} \frac{a'_d}{a_d}$$

(recall that $\frac{0}{0}$ is defined to be 0). Notice that r'_d can be ∞, but then $p'_d = 0$. Define

$$r' = \sum_{D} p'_d r'_d, \quad w' = \sum_{D} p'_d w'_d, \quad \gamma' = \sum_{D} p'_d \gamma'_d.$$

Clearly,

$$\gamma'_d(x,y) = \begin{cases} \dfrac{a_d}{a'_d} \gamma_d(x,y) & \text{for } (x,y) \in X' \times Y', \\ 0 & \text{for } (x,y) \notin X' \times Y'. \end{cases}$$

Therefore, by the definitions,

$$\gamma'(x,y) = \begin{cases} \dfrac{1}{\delta} \gamma(x,y) & \text{for } (x,y) \in X' \times Y', \\ 0 & \text{for } (x,y) \notin X' \times Y'. \end{cases}$$

So

$$\gamma' = \left(\frac{1}{\delta} \gamma(X' \times Y') \right) \gamma_{X' \times Y'}.$$

Therefore, we also have

$$\sum_D p'_d = \| \gamma' \|_1 = \gamma(X' \times Y')/\delta.$$

We would like to use induction (of the lemma), with the family $\{\gamma'_d\}_{d \in D}$, to deduce a lower bound for $w(\mu_{X' \times Y'})$. In order to use the lemma, we need bounds for r', $\| \gamma' - \mu_{X' \times Y'} \|_1$, and for w'.
First note that

$$\begin{aligned} r'_d a'_d &= \mu \left[(X'_d \times (Y' - Y'_d)) \cup ((X' - X'_d) \times Y'_d) \right] \\ &\le \mu \left[(X_d \times (Y - Y_d)) \cup ((X - X_d) \times Y_d) \right] = r_d a_d. \end{aligned}$$

Therefore, defining

$$\epsilon'_1 = \frac{\epsilon_1}{\delta}$$

we have

$$r' = \sum_D p'_d r'_d = \frac{1}{\delta} \sum_D p_d \frac{a'_d}{a_d} r'_d \le \frac{1}{\delta} \sum_D p_d r_d = \frac{r}{\delta} \le \frac{\epsilon_1}{\delta} = \epsilon'_1.$$

Also, define

$$\epsilon'_0 = \| \gamma' - \mu_{X' \times Y'} \|_1;$$

then clearly,

$$\begin{aligned} \delta \epsilon'_0 = \| \delta \gamma' - \delta \mu_{X' \times Y'} \|_1 &= \| \gamma(X' \times Y') \gamma_{X' \times Y'} - \mu(X' \times Y') \mu_{X' \times Y'} \|_1 \\ = \sum_{X' \times Y'} |\gamma(x,y) - \mu(x,y)| &\le \sum_{X \times Y} |\gamma(x,y) - \mu(x,y)| = \| \gamma - \mu \|_1 = \epsilon_0. \end{aligned}$$

Thus we have bounds for r',$\| \gamma' - \mu_{X' \times Y'} \|_1$. The bound for w' will follow from the following two claims.

CLAIM 6.2.

$$\| \gamma \|_1 - \delta \| \gamma' \|_1 \le \epsilon_0 - \delta \epsilon'_0 + (1 - \delta).$$

Proof. Define $Z = (X \times Y) - (X' \times Y')$, and $Z' = X' \times Y'$. Then $\gamma = \gamma(Z)\gamma_Z + \gamma(Z')\gamma_{Z'}$, and $\mu = \mu(Z)\mu_Z + \mu(Z')\mu_{Z'}$. Therefore,

$$\| \gamma \|_1 = \| \gamma(Z)\gamma_Z \|_1 + \| \gamma(Z')\gamma_{Z'} \|_1 = \| \gamma(Z)\gamma_Z \|_1 + \delta \| \gamma' \|_1 . \tag{4}$$

In addition,

$$\begin{aligned} \epsilon_0 &\ge \| \gamma - \mu \|_1 = \| \gamma(Z)\gamma_Z - \mu(Z)\mu_Z \|_1 + \| \gamma(Z')\gamma_{Z'} - \mu(Z')\mu_{Z'} \|_1 \\ &= \| \gamma(Z)\gamma_Z - \mu(Z)\mu_Z \|_1 + \| \delta\gamma' - \delta\mu_{Z'} \|_1 = \| \gamma(Z)\gamma_Z - \mu(Z)\mu_Z \|_1 + \delta \epsilon'_0. \end{aligned}$$

Thus, by the triangle inequality,

$$\epsilon_0 - \delta\epsilon_0' \geq \| \gamma(Z)\gamma_Z - \mu(Z)\mu_Z \|_1 \geq \| \gamma(Z)\gamma_Z \|_1 - \| \mu(Z)\mu_Z \|_1 = \| \gamma(Z)\gamma_Z \|_1 - (1-\delta),$$

so,

$$\| \gamma(Z)\gamma_Z \|_1 \leq \epsilon_0 - \delta\epsilon_0' + (1-\delta).$$

The proof follows by substituting this in (4). □

CLAIM 6.3.

$$w' \geq (w - (1-\delta) - \epsilon_0 + \delta\epsilon_0')/\delta.$$

Proof. Any protocol for the game with the measure $\mu_{X_d \times Y_d}$ defines a corresponding protocol for the game with the measure $\mu_{X_d' \times Y_d'}$. Take the best protocol for $\mu_{X_d \times Y_d}$. This protocol for $\mu_{X_d \times Y_d}$ satisfies Q on a set of points of μ-measure $w_d a_d$. At most μ-measure of $a_d - a_d' = \mu(X_d \times Y_d) - \mu(X_d' \times Y_d')$ is outside $X_d' \times Y_d'$. Therefore, at least μ-measure of $w_d a_d - (a_d - a_d')$ is inside. Thus, the corresponding protocol for $\mu_{X_d' \times Y_d'}$ satisfies Q on a set of points of μ-measure $w_d a_d - (a_d - a_d')$ (at least). But this cannot be more than $w_d' a_d'$; thus,

$$w_d' a_d' \geq w_d a_d - a_d + a_d'.$$

Therefore,

$$p_d' w_d' = p_d \frac{1}{\delta} \frac{a_d'}{a_d} w_d' \geq p_d \frac{1}{\delta} \frac{1}{a_d} [w_d a_d - a_d + a_d'] = \frac{1}{\delta} p_d w_d - \frac{1}{\delta} p_d + p_d',$$

and we have

$$\begin{aligned} w' &= \sum_D p_d' w_d' \geq \frac{1}{\delta} \sum_D p_d w_d - \frac{1}{\delta} \sum_D p_d + \sum_D p_d' \\ &= \frac{1}{\delta} w - \frac{1}{\delta} \| \gamma \|_1 + \| \gamma' \|_1 = \frac{1}{\delta} [w - (\| \gamma \|_1 - \delta \| \gamma' \|_1)], \end{aligned}$$

and by Claim 6.2

$$w' \geq (w - (1-\delta) - \epsilon_0 + \delta\epsilon_0')/\delta. \qquad \square$$

By Claim 6.3, we can conclude

$$\begin{aligned} w' - 2\sqrt{\epsilon_1'} - \epsilon_0' &\geq \frac{1}{\delta}(w - (1-\delta) - \epsilon_0 + \delta\epsilon_0') - 2\sqrt{\frac{\epsilon_1}{\delta}} - \epsilon_0' \\ &= \frac{1}{\delta}(w - (1-\delta) - 2\sqrt{\delta}\sqrt{\epsilon_1} - \epsilon_0) \geq \frac{1}{\delta}(w - (1-\delta) - 2\sqrt{\epsilon_1} - \epsilon_0). \end{aligned}$$

Now, we can use induction and apply the lemma for $\mu_{X' \times Y'}$, with the family $\{\gamma_d'\}_{d \in D}$ to get a lower bound

$$w(\mu_{X' \times Y'}) \geq f\left[w' - 2\sqrt{\epsilon_1'} - \epsilon_0'\right] \geq f\left[\frac{1}{\delta}(w - 2\sqrt{\epsilon_1} - \epsilon_0 - (1-\delta))\right]$$

(since f is monotone). Define $z = w - 2\sqrt{\epsilon_1} - \epsilon_0$ to get

$$w(\mu_{X' \times Y'}) \geq f\left[\frac{1}{\delta}(z - (1-\delta))\right].$$

Recall that $w_{d_0} \geq w - \sqrt{\epsilon_1} - \epsilon_0$. Since $a_{d_0} + r_{d_0} a_{d_0} + \delta = \mu(X \times Y) = 1$, we have $a_{d_0} = (1-\delta)/(1+r_{d_0})$. But $r_{d_0} \leq \sqrt{\epsilon_1}$, and therefore

$$a_{d_0} \geq (1-\delta)/(1+\sqrt{\epsilon_1}) \geq (1-\delta)(1-\sqrt{\epsilon_1})$$

and we have

$$a_{d_0} w_{d_0} \geq (1-\delta)(1-\sqrt{\epsilon_1})(w - \sqrt{\epsilon_1} - \epsilon_0) \geq (1-\delta)(w - 2\sqrt{\epsilon_1} - \epsilon_0) = (1-\delta)z.$$

Recall that $X_{d_0} \cap X' = \emptyset$, and $Y_{d_0} \cap Y' = \emptyset$. Given the best protocol for $\mu_{X_{d_0} \times Y_{d_0}}$ and the best protocol for $\mu_{X' \times Y'}$, define a protocol for μ that behaves like the first one on $X_{d_0} \times Y_{d_0}$ and like the other one on $X' \times Y'$. This protocol satisfies Q on a set of points of μ-measure $\mu(X_{d_0} \times Y_{d_0}) w(\mu_{X_{d_0} \times Y_{d_0}}) + \mu(X' \times Y') w(\mu_{X' \times Y'})$ (at least), and therefore proves that

$$\begin{aligned} w(\mu) &\geq a_{d_0} w_{d_0} + \delta w(\mu_{X' \times Y'}) \\ &\geq (1-\delta)z + \delta f\left[\frac{1}{\delta}(z - (1-\delta))\right] \geq f(z) = f(w - 2\sqrt{\epsilon_1} - \epsilon_0), \end{aligned}$$

which proves the lemma. The assumption $z \leq 0 \Rightarrow f(z) \leq 0$ is needed for the base case of the induction. We remark that if $\delta = 0$ the lemma is proved simply from Claim 6.1 or by a continuity argument. □

COROLLARY 6.2. *Under the assumptions of Lemma* 6.1*,*

$$w(\mu) \geq f(w - 2\sqrt{\epsilon_1} - \epsilon_0),$$

where f is defined by

$$f(z) = \begin{cases} \frac{1}{2}z^2 & \text{for } 0 \leq z \leq 1, \\ 0 & \text{for } z \leq 0. \end{cases}$$

Proof. We just have to prove the inequality (as in Lemma 6.1) for f.

Case a. $z \geq 1 - \delta$.

In this case

$$\begin{aligned} (1-\delta)z + \delta f\left[\frac{1}{\delta}(z-(1-\delta))\right] - f(z) &= (1-\delta)z + \frac{1}{2\delta}(z-(1-\delta))^2 - \frac{1}{2}z^2 \\ &= \frac{1}{2\delta}\left[2\delta(1-\delta)z + z^2 - 2(1-\delta)z + (1-\delta)^2 - \delta z^2\right] \\ &= \frac{1}{2\delta}\left[z^2(1-\delta) - 2z(1-\delta)^2 + (1-\delta)^2\right] \\ &\geq \frac{1-\delta}{2\delta}\left[z^2 - 2z(1-\delta) + (1-\delta)^2\right] \\ &= \frac{1-\delta}{2\delta}\left[z-(1-\delta)\right]^2 \quad \geq \quad 0. \end{aligned}$$

Case b. $z \leq 1 - \delta$.

In this case

$$\begin{aligned} (1-\delta)z + \delta f\left[\frac{1}{\delta}(z-(1-\delta))\right] - f(z) &= (1-\delta)z + 0 - \frac{1}{2}z^2 \\ &\geq z^2 - \frac{1}{2}z^2 = \frac{1}{2}z^2 \geq 0. \end{aligned}$$ □

It is sometimes convenient to denote $v(\mu) = 1 - w(\mu)$, $v = 1 - w$, and $g(t) = 1 - f(1 - t)$. In these notations the inequality

$$(1-\delta)z + \delta f\left[\frac{1}{\delta}(z - (1-\delta))\right] - f(z) \geq 0$$

(for $0 \leq z \leq 1$, $0 < \delta \leq 1$), is equivalent (by setting $t = 1 - z$) to

$$(1-\delta)t + \delta g(t/\delta) \leq g(t)$$

(for $0 \leq t \leq 1$, $0 < \delta \leq 1$).

Corollary 6.3. *Under the assumptions of Lemma* 6.1*, if $g : \mathbf{R} \to \mathbf{R}$ is an increasing monotone function such that $g(t) \geq 1$, for $t \geq 1$, and such that for all $0 \leq t \leq 1$, $0 < \delta \leq 1$, we have*

$$(1-\delta)t + \delta g(t/\delta) - g(t) \leq 0;$$

then,

$$v(\mu) \leq g(v + 2\sqrt{\epsilon_1} + \epsilon_0).$$

Corollary 6.4. *Under the assumptions of Lemma* 6.1*,*

$$v(\mu) \leq 2\sqrt{v + 2\sqrt{\epsilon_1} + \epsilon_0}.$$

Proof. Take $g(t) = 2\sqrt{t}$, then for $0 \leq t \leq 1$, $0 < \delta \leq 1$,

$$\begin{aligned}(1-\delta)t + \delta g(t/\delta) - g(t) &= (1-\delta)t + 2\sqrt{\delta}\sqrt{t} - 2\sqrt{t}\\ &= (1+\sqrt{\delta})(1-\sqrt{\delta})\sqrt{t}\sqrt{t} - 2(1-\sqrt{\delta})\sqrt{t}\\ &= (1-\sqrt{\delta})\sqrt{t}[(1+\sqrt{\delta})\sqrt{t} - 2] \quad \leq \quad 0\end{aligned}$$

(because $(1+\sqrt{\delta})\sqrt{t} - 2 \leq 2 - 2 = 0$). □

It will be convenient to define the function $W_1 : \mathbf{R} \to \mathbf{R}$ in the following way:

$$W_1(z) = \sup_f f(z)$$

where the supremum is taken over all the strictly increasing monotone functions, satisfying the conditions as in Lemma 6.1. Conversely, define $W_2 : \mathbf{R} \to \mathbf{R}$ by

$$W_2(z) = \inf_f f^{-1}(z)$$

where the infimum is taken over the same family of functions.

Take, for convenience, $\epsilon_0 = \epsilon_1 = \epsilon$, and assume for convenience that $w \leq 1$. Lemma 6.1 can now be restated in the following way.

Corollary 6.5. *If for a family $\{\gamma_d\}_{d \in D}$ we have $\| \gamma - \mu \|_1 \leq \epsilon$ and $r \leq \epsilon$, then*

$$w(\mu) \geq W_1(w - O(\sqrt{\epsilon}))$$

or, conversely,

$$w \leq W_2(w(\mu)) + O(\sqrt{\epsilon}).$$

By Corollaries 6.2 and 6.4 we have

$$0 < z \Rightarrow 0 < W_1(z), \quad \lim_{z\to 1} W_1(z) = 1, \quad \lim_{z \to 0} W_2(z) = 0, \quad z < 1 \Rightarrow W_2(z) < 1$$

(where the last fact is the important one for us). Also, since f was monotone, $W_1(z), W_2(z)$ are both monotone. Thus, W_1, W_2 are "well behaved." In this manuscript we will not investigate their exact behavior.

6.2. Characterization of measures $\boldsymbol{\vartheta}$, with small $\mathbf{V}\left(\boldsymbol{\vartheta} \parallel \boldsymbol{\mu}\right)$. Lemma 4.3 talks about measures γ_d, with small $\mathbf{V_X}\left(\gamma_d \parallel \mu\right)$ and small $\mathbf{V_Y}\left(\gamma_d \parallel \mu\right)$. We would like to prove this lemma by a reduction to Lemma 6.1. In order to do that, we need one more lemma that gives a characterization of measures ϑ, with small $\mathbf{V_X}\left(\vartheta \parallel \mu\right)$ and small $\mathbf{V_Y}\left(\vartheta \parallel \mu\right)$. Again, $\mu : X \times Y \to \mathbf{R}^+$ and $\vartheta : X \times Y \to \mathbf{R}^+$ are probability measures.

LEMMA 6.6. *If $\mathbf{V_X}\left(\vartheta \parallel \mu\right), \mathbf{V_Y}\left(\vartheta \parallel \mu\right) \le \epsilon$ then, for some m, there exist partitions $X = \bigcup_{i=1}^m X_i$, $Y = \bigcup_{i=1}^m Y_i$, and positive weights $\{q_i\}_{i=1}^m$ (i.e., for all $i:\ q_i \ge 0$), such that the function $h : X \times Y \to \mathbf{R}$, defined by*

$$h(x,y) = \sum_{i=1}^m q_i \mu_{X_i \times Y_i}(x,y)$$

satisfies

$$\parallel \vartheta - h \parallel_1 \le O(\epsilon^{1/8})$$

and such that

$$\sum_{i=1}^m q_i r_i \le O(\epsilon^{1/8}),$$

where

$$r_i = \frac{\mu\left[(X_i \times (Y - Y_i)) \cup ((X - X_i) \times Y_i)\right]}{\mu(X_i \times Y_i)}.$$

Proof. Assume that $\vartheta(x), \vartheta(y), \mu(x), \mu(y)$ take only strictly positive values (otherwise, just add small constants and normalize). This is done only to simplify the notations. Assume that ϵ is small enough ($\epsilon < 2^{-10}$ is enough), (for $\epsilon \ge 2^{-10}$, $O(\epsilon^{1/8}) = O(1)$, thus $h = \mu$ does the job).

The first claim describes the basic structure of the measure ϑ.

CLAIM 6.4.

$$\left\| \vartheta(x,y) - \frac{\vartheta(x)}{\mu(x)}\mu(x,y) \right\|_1 \le \sqrt{2\ln 2}\sqrt{\epsilon},$$

$$\left\| \vartheta(x,y) - \frac{\vartheta(y)}{\mu(y)}\mu(x,y) \right\|_1 \le \sqrt{2\ln 2}\sqrt{\epsilon}.$$

Proof. We will prove the first inequality. The second can be proved in a similar manner.

By Lemma 3.4 we have

$$\begin{aligned}\sum_{x\in X} \vartheta(x)\left(\parallel \vartheta(x,\cdot) - \mu(x,\cdot) \parallel_1\right)^2 &\le (2\ln 2)\sum_{x\in X}\vartheta(x)\mathbf{D}\left(\vartheta(x,\cdot) \parallel \mu(x,\cdot)\right)\\ &= (2\ln 2)\mathbf{V_X}\left(\vartheta \parallel \mu\right) \ \le\ (2\ln 2)\epsilon.\end{aligned}$$

Since for any random variable z: $(\mathbf{E}(z))^2 \le \mathbf{E}(z^2)$, we have

$$\sum_{x\in X} \vartheta(x)\left(\parallel \vartheta(x,\cdot) - \mu(x,\cdot) \parallel_1\right) \ \le\ \sqrt{2\ln 2}\sqrt{\epsilon},$$

but

$$\sum_{x\in X}\vartheta(x)\left(\|\,\vartheta(x,\cdot)-\mu(x,\cdot)\,\|_1\right) \;=\; \sum_{x\in X}\vartheta(x)\sum_{y\in Y}\left|\frac{\vartheta(x,y)}{\vartheta(x)}-\frac{\mu(x,y)}{\mu(x)}\right|$$

$$=\;\sum_{X\times Y}\left|\vartheta(x,y)-\frac{\vartheta(x)}{\mu(x)}\mu(x,y)\right| \;=\; \left\|\,\vartheta(x,y)-\frac{\vartheta(x)}{\mu(x)}\mu(x,y)\,\right\|_1$$

and the claim follows. □

Define

$$R_X(x)=\frac{\vartheta(x)}{\mu(x)},\quad R_Y(y)=\frac{\vartheta(y)}{\mu(y)},\quad R(x,y)=\frac{R_Y(y)}{R_X(x)}$$

(note that $R(x,y)$ is asymmetric). By our assumption, these values are always well defined and strictly positive. Define

$$h_1(x,y)=R_X(x)\mu(x,y),\quad h_2(x,y)=R_Y(y)\mu(x,y).$$

We proved

$$\|\,\vartheta-h_1\,\|_1\le\sqrt{2\ln 2}\sqrt{\epsilon},\quad \|\,\vartheta-h_2\,\|_1\le\sqrt{2\ln 2}\sqrt{\epsilon}.$$

Therefore, by the triangle inequality we also have

$$\|\,h_1\,\|_1\le\|\,\vartheta\,\|_1+\|\,\vartheta-h_1\,\|_1\le\;1+\sqrt{2\ln 2}\sqrt{\epsilon}.$$

Similarly,

$$\|\,h_2\,\|_1\le\;1+\sqrt{2\ln 2}\sqrt{\epsilon}$$

and

$$\|\,h_1-h_2\,\|_1\le\|\,\vartheta-h_1\,\|_1+\|\,\vartheta-h_2\,\|_1\le\;2\sqrt{2\ln 2}\sqrt{\epsilon}.$$

The last inequality shows that most of the measure μ is concentrated on pairs (x,y), with $R_X(x)$ very close to $R_Y(y)$. In the simplest case, where $\epsilon=0$, the entire measure μ is concentrated on pairs, with $R_X(x)=R_Y(y)$. In this case we can partition X according to $R_X(x)$, and Y according to $R_Y(y)$, and define the weight of a subset as $R_X(x)$ (or $R_Y(y)$). The lemma follows then from Claim 6.4 and from the last inequality. In the general case the intuition will be the same, but we will have to partition X and Y into subsets, according to the value of $R_X(x)$ and $R_Y(y)$, where each subset allows small deviations in the value.

Denote

$$\epsilon_1=2\sqrt{2\ln 2}\sqrt{\epsilon}$$

and define

$$Z=\left\{(x,y)\in X\times Y\;\;|\;\;|1-R(x,y)|\ge\sqrt{\epsilon_1}\right\}.$$

Claim 6.5.

$$\sum_Z h_1(x,y)\le\sqrt{\epsilon_1},\quad \sum_Z h_2(x,y)\le\sqrt{\epsilon_1}.$$

Proof. By the definitions

$$\| \, h_1(x,y)\,(1-R(x,y))\, \|_1 = \| \, h_1 - h_2 \, \|_1 \le \; 2\sqrt{2\ln 2}\sqrt{\epsilon} \; = \; \epsilon_1.$$

Therefore,

$$\sum_Z h_1(x,y)\sqrt{\epsilon_1} \; \le \; \sum_Z h_1(x,y)\,|1-R(x,y)| \; \le \; \sum_{X,Y} h_1(x,y)\,|1-R(x,y)| \; \le \; \epsilon_1.$$

Thus,

$$\sum_Z h_1(x,y) \le \sqrt{\epsilon_1}.$$

The second inequality is proved in the same way. □

Denote

$$\epsilon_2 = \sqrt{-\ln(1-\sqrt{\epsilon_1})}.$$

Then since $\epsilon < 2^{-10}$,

$$\epsilon_2 \le \sqrt{\ln(1+2\sqrt{\epsilon_1})} \le \sqrt{2\sqrt{\epsilon_1}} \le 2\epsilon^{1/8}.$$

For $(x,y) \notin Z$, we have

$$1-\sqrt{\epsilon_1} \le R(x,y) \le 1+\sqrt{\epsilon_1}.$$

Hence,

$$\ln(1-\sqrt{\epsilon_1}) \le \ln R(x,y) \le \ln(1+\sqrt{\epsilon_1}) \le -\ln(1-\sqrt{\epsilon_1}).$$

Thus, $(x,y) \notin Z$ implies

$$-\epsilon_2^2 \le \ln R(x,y) \le \epsilon_2^2.$$

Let r be a random variable uniformly distributed in the interval $[0,1]$. For every integer i define

$$X_i(r) = \{x \in X \;\mid\; (i-1+r)\epsilon_2 \le \ln R_X(x) < (i+r)\epsilon_2\},$$
$$Y_i(r) = \{y \in Y \;\mid\; (i-1+r)\epsilon_2 \le \ln R_Y(y) < (i+r)\epsilon_2\}.$$

Then, for all r, $\{X_i(r)\}_{i=-\infty}^{\infty}$ is a partition of X, and $\{Y_i(r)\}_{i=-\infty}^{\infty}$ is a partition of Y. Define

$$\hat{Z}(r) = \left\{ (x,y) \in X \times Y \;\middle|\; (x,y) \notin \bigcup_{i=-\infty}^{\infty} X_i \times Y_i \right\}.$$

We will prove that for some r these partitions satisfy the lemma.

Claim 6.6.

$$(x,y) \notin Z \;\; \Rightarrow \;\; \Pr_r\left[(x,y) \in \hat{Z}(r)\right] \le \epsilon_2.$$

Proof. $(x,y) \notin Z$ implies

$$|\ln R_X(x) - \ln R_Y(y)| = |\ln R(x,y)| \le \epsilon_2^2.$$

Assume w.l.o.g. that $\ln R_X(x) \ge \ln R_Y(y)$. For a fixed r, $(x,y) \in X_i(r) \times Y_i(r)$ for some i, unless there exists in the interval $[\ln R_Y(y), \ln R_X(x)]$, a number of the form $(j+r)\epsilon_2$ (for some integer j). The probability for that (over r) is

$$\frac{\ln R_X(x) - \ln R_Y(y)}{\epsilon_2} \le \frac{\epsilon_2^2}{\epsilon_2} = \epsilon_2. \qquad \square$$

CLAIM 6.7.

$$\mathbf{E}_r\left(\sum_{\hat{Z}(r)} h_1(x,y)\right) \le 3\epsilon_2, \quad \mathbf{E}_r\left(\sum_{\hat{Z}(r)} h_2(x,y)\right) \le 3\epsilon_2.$$

Proof. By changing the order of the summations,

$$\begin{aligned}\mathbf{E}_r\left(\sum_{\hat{Z}(r)} h_1(x,y)\right) &= \sum_{X\times Y} h_1(x,y)\mathrm{Pr}_r\left[(x,y)\in\hat{Z}(r)\right]\\ &= \sum_{X\times Y-Z} h_1(x,y)\mathrm{Pr}_r\left[(x,y)\in\hat{Z}(r)\right] + \sum_Z h_1(x,y)\mathrm{Pr}_r\left[(x,y)\in\hat{Z}(r)\right]\\ &\le \sum_{X\times Y-Z} h_1(x,y)\epsilon_2 + \sum_Z h_1(x,y)1 \le \epsilon_2\sum_{X\times Y} h_1(x,y) + \sum_Z h_1(x,y)\\ &\le \epsilon_2 \parallel h_1 \parallel_1 + \sqrt{\epsilon_1} \le \epsilon_2(1+\sqrt{2\ln 2}\sqrt{\epsilon}) + \sqrt{\epsilon_1} \le 3\epsilon_2.\end{aligned}$$

The second inequality is proved in the same way. $\square$

Since $h_1(x,y), h_2(x,y)$ are always positive, we can conclude from the last claim the existence of r_0 with

$$\sum_{\hat{Z}(r)} h_1(x,y) \le 6\epsilon_2, \quad \sum_{\hat{Z}(r)} h_2(x,y) \le 6\epsilon_2.$$

Fix this r_0. Define

$$X_i = X_i(r_0), \quad Y_i = Y_i(r_0), \quad \hat{Z} = \hat{Z}(r_0).$$

We will prove that these partitions satisfy the lemma. Define $h_3 : X \times Y \to \mathbf{R}$ by

$$h_3(x,y) = \begin{cases} h_1(x,y) & \text{for } (x,y)\notin\hat{Z},\\ 0 & \text{for } (x,y)\in\hat{Z}.\end{cases}$$

Then clearly,

$$\parallel h_1 - h_3 \parallel_1 = \sum_{X,Y} |h_1(x,y) - h_3(x,y)| = \sum_{\hat{Z}} h_1(x,y) \le 6\epsilon_2.$$

By the definitions,

$$h_3(x,y) = \sum_{i=-\infty}^{\infty} R_X(x)\mu(X_i \times Y_i)\mu_{X_i\times Y_i}(x,y).$$

Define

$$q_i = e^{(i-1+r_0)\epsilon_2}\mu(X_i \times Y_i)$$

and

$$h(x,y) = \sum_{i=-\infty}^{\infty} q_i \mu_{X_i \times Y_i}(x,y).$$

CLAIM 6.8.

$$\| h_3 - h \|_1 \leq 2\epsilon_2.$$

Proof. For $x \in X_i$,

$$e^{(i-1+r_0)\epsilon_2} \leq R_X(x) < e^{(i+r_0)\epsilon_2};$$

thus, by the definition of q_i,

$$R_X(x)\mu(X_i \times Y_i) \;\; \geq \;\; q_i \;\; > \;\; R_X(x)\mu(X_i \times Y_i)e^{-\epsilon_2} \geq R_X(x)\mu(X_i \times Y_i)(1-\epsilon_2).$$

Therefore, for $x \in X_i$

$$0 \;\leq\; R_X(x)\mu(X_i \times Y_i) - q_i \;\leq\; R_X(x)\mu(X_i \times Y_i)\epsilon_2.$$

Since $\mu_{X_i \times Y_i}(x,y) > 0 \Rightarrow x \in X_i$, we have for all x, y,

$$|R_X(x)\mu(X_i \times Y_i) - q_i|\, \mu_{X_i \times Y_i}(x,y) \;\leq\; \epsilon_2 R_X(x)\mu(X_i \times Y_i)\mu_{X_i \times Y_i}(x,y)$$

and therefore,

$$\begin{aligned}
\| h_3 - h \|_1 &= \left\| \sum_{i=-\infty}^{\infty} R_X(x)\mu(X_i \times Y_i)\mu_{X_i \times Y_i}(x,y) - \sum_{i=-\infty}^{\infty} q_i \mu_{X_i \times Y_i}(x,y) \right\|_1 \\
&\leq \left\| \sum_{i=-\infty}^{\infty} |R_X(x)\mu(X_i \times Y_i) - q_i|\, \mu_{X_i \times Y_i}(x,y) \right\|_1 \\
&\leq \left\| \sum_{i=-\infty}^{\infty} \epsilon_2 R_X(x)\mu(X_i \times Y_i)\mu_{X_i \times Y_i}(x,y) \right\|_1 \\
&= \epsilon_2 \| h_3 \|_1 \leq\; \epsilon_2 \| h_1 \|_1 \;\leq\; 2\epsilon_2. \qquad \square
\end{aligned}$$

By the triangle inequality, we can now conclude that

$$\begin{aligned}
\| h - \vartheta \|_1 &\leq \| \vartheta - h_1 \|_1 + \| h_1 - h_3 \|_1 + \| h_3 - h \|_1 \\
&\leq \sqrt{2\ln 2}\sqrt{\epsilon} + 3\epsilon_2 + 2\epsilon_2 = O(\epsilon^{1/8}).
\end{aligned}$$

Notice that in the definition of h, the sum is actually finite, because X, Y are finite ($\mu(X_i \times Y_i)$ can take a nonzero value only a finite number of times).

CLAIM 6.9.

$$\sum_{i=-\infty}^{\infty} q_i \frac{\mu[X_i \times (Y - Y_i)]}{\mu(X_i \times Y_i)} \leq O(\epsilon^{1/8}),$$

$$\sum_{i=-\infty}^{\infty} q_i \frac{\mu[(X - X_i) \times Y_i]}{\mu(X_i \times Y_i)} \leq O(\epsilon^{1/8}).$$

Proof. Notice that $\mu(X_i \times Y_i)$ can be 0, only if q_i is also 0. As in the previous claim for $x \in X_i$, $q_i \le R_X(x)\mu(X_i \times Y_i)$. Thus,

$$q_i \frac{\mu[X_i \times (Y - Y_i)]}{\mu(X_i \times Y_i)} = \sum_{(x,y)\in X_i \times (Y-Y_i)} \frac{q_i}{\mu(X_i \times Y_i)} \mu(x,y) \le \sum_{(x,y)\in X_i \times (Y-Y_i)} R_X(x)\mu(x,y),$$

and therefore,

$$\sum_{i=-\infty}^{\infty} q_i \frac{\mu[X_i \times (Y - Y_i)]}{\mu(X_i \times Y_i)} \le \sum_{i=-\infty}^{\infty} \left(\sum_{(x,y)\in X_i \times (Y-Y_i)} R_X(x)\mu(x,y) \right)$$
$$= \sum_{(x,y)\in \hat{Z}} R_X(x)\mu(x,y) = \sum_{\hat{Z}} h_1(x,y) \le 6\epsilon_2 = O(\epsilon^{1/8}).$$

The second inequality is proved in the same way. □

Claim 6.9 and the inequality before give the proof of Lemma 6.6. □

6.3. Proof of Lemma 4.3. Lemma 4.3 will follow as a simple application of Lemmas 6.1 and 6.6. Given ϑ, with

$$\mathbf{V_X}\left(\vartheta \parallel \mu\right), \mathbf{V_Y}\left(\vartheta \parallel \mu\right) \le \epsilon$$

take m, $\{q_i\}_{i=1}^m$, $\{X_i\}_{i=1}^m$, $\{Y_i\}_{i=1}^m$ from Lemma 6.6. Define

$$\gamma_i = \mu_{X_i \times Y_i}, \quad w_i = w(\gamma_i)$$

and

$$r_i = \frac{\mu[(X_i \times (Y - Y_i)) \cup ((X - X_i) \times Y_i)]}{\mu(X_i \times Y_i)}$$

as in Lemma 6.1. By Lemma 6.6 we have

$$\sum_{i=1}^{m} q_i r_i \le O(\epsilon^{1/8})$$

and

$$\left\| \vartheta - \sum_{i=1}^{m} q_i \gamma_i \right\|_1 \le O(\epsilon^{1/8}).$$

Every protocol for ϑ defines also a protocol for each one of the γ_i-s. Therefore, the last inequality also gives

$$w(\vartheta) \le \sum_{i=1}^{m} q_i w_i + O(\epsilon^{1/8}),$$

which reflects the fact that the function $w(\gamma)$ is concave and has a Lipschitz constant of 1 (see the Introduction).

Lemma 4.3 can be proved now in the following way. Given a family $\{\gamma_d\}_{d\in D}$ of probability measures on $X \times Y$ and weights $\{p_d\}_{d\in D}$, such that $\sum_{d\in D} p_d = 1$, and such that for all $d: p_d \ge 0$, define

$$w_d = w(\gamma_d), \quad w = \sum_{d\in D} p_d w_d, \quad V_d = \mathbf{V}\left(\gamma_d \parallel \mu\right).$$

Lemma 4.3 assumes

$$\sum_{d\in D} p_d V_d \;\le\; \epsilon$$

and

$$\mathbf{D}\left(\sum_{d\in D} p_d \gamma_d \,\middle\|\, \mu\right) \;\le\; \epsilon,$$

and therefore, by Lemma 3.4

$$\left\| \mu - \sum_{d\in D} p_d \gamma_d \right\|_1 \;\le\; O(\epsilon^{1/2}).$$

For each measure γ_d, we have by the previous discussion, $\{\gamma_{d,i}\}_{i=1}^{m_d}$, and $\{q_{d,i}\}_{i=1}^{m_d}$, such that

$$\sum_{i=1}^{m_d} q_{d,i} r_{d,i} \;\le\; O(V_d^{1/8})$$

and

$$\left\| \gamma_d - \sum_{i=1}^{m_d} q_{d,i} \gamma_{d,i} \right\|_1 \;\le\; O(V_d^{1/8});$$

and as before, we also have

$$w_d \;\le\; \sum_{i=1}^{m_d} q_{d,i} w_{d,i} + O(V_d^{1/8})$$

(where $r_{d,i}, w_{d,i}$ are defined as before). Define the set $\hat{D}$ by

$$\hat{D} = \{(d,i) \;\mid\; d \in D \;,\; 1 \le i \le m_d\}.$$

For each $\hat{d} = (d,i) \in \hat{D}$, define the weight

$$\hat{p}_{(d,i)} = p_d q_{d,i}$$

and the measure

$$\hat{\gamma}_{(d,i)} = \gamma_{d,i}.$$

Look at the family of measures $\{\hat{\gamma}_{(d,i)}\}_{(d,i)\in\hat{D}}$. For this family, define as before

$$\hat{w}_{(d,i)} = w(\hat{\gamma}_{(d,i)}), \quad \hat{w} = \sum_{\hat{d}\in\hat{D}} \hat{p}_{\hat{d}} \hat{w}_{\hat{d}}, \quad \hat{r}_{(d,i)} = r_{d,i}.$$

Then, by the convexity of the function $f(z) = z^{1/8}$, we have

$$\sum_{\hat{d}\in\hat{D}} \hat{p}_{\hat{d}} \hat{r}_{\hat{d}} = \sum_{d\in D} p_d \sum_{i=1}^{m_d} q_{d,i} r_{d,i} \;\;\le\;\; \sum_{d\in D} p_d O(V_d^{1/8}) \;\;\le\;\; O\left(\sum_{d\in D} p_d V_d\right)^{1/8}$$
$$\le O(\epsilon^{1/8})$$

and

$$\begin{aligned}\left\| \mu - \sum_{\hat{d}\in\hat{D}} \hat{p}_{\hat{d}}\hat{\gamma}_{\hat{d}} \right\|_1 &\leq \left\| \mu - \sum_{d\in D} p_d\gamma_d \right\|_1 + \left\| \sum_{d\in D} p_d\gamma_d - \sum_{\hat{d}\in\hat{D}} \hat{p}_{\hat{d}}\hat{\gamma}_{\hat{d}} \right\|_1 \\ &\leq O(\epsilon^{1/2}) + \sum_{d\in D} p_d \left\| \gamma_d - \sum_{i=1}^{m_d} q_{d,i}\gamma_{d,i} \right\|_1 \\ &\leq O(\epsilon^{1/2}) + \sum_{d\in D} p_d O(V_d^{1/8}) \leq O(\epsilon^{1/2}) + O\left(\sum_{d\in D} p_d V_d\right)^{1/8} \\ &\leq O(\epsilon^{1/8}),\end{aligned}$$

and also as before,

$$\begin{aligned}\hat{w} = \sum_{\hat{d}\in\hat{D}} \hat{p}_{\hat{d}}\hat{w}_{\hat{d}} &= \sum_{d\in D} p_d \sum_{i=1}^{m_d} q_{d,i}w_{d,i} \geq \sum_{d\in D} p_d \left(w_d - O(V_d^{1/8})\right) \\ &= w - \sum_{d\in D} p_d O(V_d^{1/8}) \\ &\geq w - O(\epsilon^{1/8}).\end{aligned}$$

Now we can apply Corollary 6.5 for the family $\{\hat{\gamma}_{\hat{d}}\}_{\hat{d}\in\hat{D}}$ to get

$$\hat{w} \leq W_2(w(\mu)) + O(\epsilon^{1/16})$$

and we can conclude

$$w \leq \hat{w} + O(\epsilon^{1/8}) \leq W_2(w(\mu)) + O(\epsilon^{1/16}).$$

7. Conclusions.

Theorem 1.1 can be generalized using methods introduced herein in many ways. Let us briefly describe two generalizations that seem to follow. The proofs were not verified as carefully as the rest of the paper and should be trusted accordingly. Several other generalizations are described in [42].

7.1. Product of games.

Given a game G and a game G', the product game $G \otimes G'$ is defined in the same manner as $G \otimes G$, i.e., if G consists of X, Y, U, V, μ, Q, and G' consists of X', Y', U', V', μ', Q', then the game $G \otimes G'$ consists of the sets $X \times X'$, $Y \times Y'$, $U \times U'$, $V \times V'$ with the measure

$$\mu \otimes \mu'((x, x'), (y, y')) = \mu(x, y)\mu'(x', y')$$

and the predicate

$$Q \otimes Q'((x, x'), (y, y'), (u, u'), (v, v')) = Q(x, y, u, v)Q'(x', y', u', v').$$

In the same way, given k games $G_1, \ldots, G_k$, the product $G_1 \otimes \cdots \otimes G_k$ is defined.

Since in the entire proof of Theorem 1.1 we didn't use the fact that the same game, G, is repeated, and since the function W from Theorem 1.1 is global (and in particular doesn't depend on the game G), the following generalization of Theorem 1.1 follows.

THEOREM 7.1. *Let $W : [0,1] \rightarrow [0,1]$ be the function from Theorem* 1.1. *Given k games $G_1, \dots, G_k$, define*

$$w = MAX[w(G_1), \dots, w(G_k)],$$

and

$$s = MAX[s(G_1), \dots, s(G_k), 2].$$

Then

$$w(G_1 \otimes \cdots \otimes G_k) \leq W(w)^{k/\log_2(s)}.$$

As before, $w = \mathrm{MAX}[s(G_1), \dots, s(G_k), 2]$ can be replaced with

$$w = \mathrm{MAX}[CC(G_1), \dots, CC(G_k), 2]$$

or with

$$w = \mathrm{MAX}[\rho(G_1), \dots, \rho(G_k), 2].$$

7.2. Probabilistic predicates. Our second generalization deals with the probabilistic case, where the predicate Q, of a game G, depends also on the random string r (i.e., Q is probabilistic). Without loss of generality we can assume that there exists a second random string, $\tilde{r}$, such that $\tilde{r}$ is independent of r (and, therefore, also of (x, y, u, v)), and such that the predicate Q depends on $x, y, u, v, \tilde{r}$ (and not on the random string r).

As before, the value of a protocol for the game G is defined to be the probability that $Q(x, y, u(x), v(y), \tilde{r}) = 1$, where (x, y) is chosen according to μ. As before, the value of the game, $w(G)$, is defined to be the maximal value of all protocols for G.

THEOREM 7.2. *Let $W : [0,1] \rightarrow [0,1]$ be the function from Theorem* 1.1. *Given a probabilistic game G (as above), with value $w(G)$, and answer-size $s(G) \geq 2$:*

$$w(G^{\otimes k}) \leq W(w(G))^{k/\log_2(s(G))}.$$

As before, $\log_2(s(G))$ can be replaced with $CC(G)$, or with $\rho(G)$, defined in the following way: Define $G_{\tilde{r}}$ to be the deterministic game obtained by fixing the second random string to $\tilde{r}$. Then define $CC(G)$ to be the maximum, taken over $\tilde{r}$, of $CC(G_{\tilde{r}})$, and define $\rho(G)$ to be the maximum, taken over $\tilde{r}$, of $\rho(G_{\tilde{r}})$.

Sketch of proof. First we claim that the proof of Theorem 1.2 holds (with minor changes) for probabilistic games as well. In section 4, the fact that the function w_l is concave is used to prove inequality (3) (and the inequality before). The concavity of the value function of a game is proved in the introduction for deterministic games. The same argument, however, holds for probabilistic games. Using this fact, one can verify that the entire argument of section 4 holds for probabilistic games as well. Now, section 4 uses Lemmas 4.2 and 4.3. Lemma 4.2 does not depend on the game G at all. The proof of Lemma 4.3 (in section 6) is based on Lemmas 6.1 and 6.6. Lemma 6.6 does not depend on the game G as well. Therefore, we just have to verify that the proof of Lemma 6.1 holds for probabilistic games. In the proof of Lemma 6.1 we use the fact that the game is deterministic only in the proof of Claim 6.3. It is not hard to see, however, that a probabilistic version of that proof can be given.

Theorem 7.2 is now proved using Theorem 1.2 in the same manner as before (see section 2). The difference is that now, given $z = (x', y', u', v') \in Z$, $\bar{Q}^1$ is still not determined on the set $A(z)$, because $\bar{Q}^1$ depends also on the second random string corresponding to the first coordinate (denote this random string by $\tilde{r}^1$). In the deterministic case, we disregarded sets $A(z)$ for every z s.t. Q is not satisfied on z. In order to be able to do the same in the probabilistic case, we will have to have a copy of $A(z)$ for every possible $\tilde{r}^1$ (for $\tilde{r}^1 = r'$ denote this copy by $A_{r'}(z)$). We then disregard every copy $A_{r'}(z)$ for every z, r' s.t. Q is not satisfied on z, r'.

A different way to see how Theorem 7.2 is proved using Theorem 1.2 is to define $q(z)$ to be the probability that $Q(z, \tilde{r}) = 1$. (For a deterministic game, $q(z)$ is always 0 or 1). It is not too hard to see that the entire argument of section 2 can be generalized to the case where $0 \leq q(z) \leq 1$.

Alternatively, we can present the proof of Theorem 7.2 in the following way.

We can view a probabilistic game G in the following equivalent way. Player I receives (as an input) the pair $(x, \tilde{r})$ and Player II receives the pair $(y, \tilde{r})$. The protocols of the players are restricted as to depend only on the first input, i.e., x for the first player, and, respectively, y for the second player, (and not on the second input $\tilde{r}$). We call such a protocol a restricted protocol.

The game G is now deterministic because we can think of the predicate as being depended only on the inputs and the answers of the two players. The class of allowed protocols, however, is now a subclass of all possible protocols. We claim that the entire proof of Theorem 1.1 (including the proof of Theorem 1.2) is correct even if we allow only restricted protocols.

In some parts of the proof (e.g., section 4), we start from a protocol, P, for a game, and obtain from this protocol many protocols for many other games. These protocols are obtained either by projection (on one coordinate) or by restriction to a product subset. In other parts (e.g., section 6) a protocol P for a game is composed from other protocols for different games. In order to see that the proof of Theorem 1.1 holds even if we allow only restricted protocols, one should verify that if the original protocols are restricted then every protocol obtained by one of these three methods is also restricted. If the new protocol is obtained by projection of a restricted protocol or by composition of restricted protocols (i.e., by the first or third method), then it is very easy to see that the new protocol is also restricted. If the new protocol is obtained by a restriction of a restricted protocol to a subset, then the new protocol is not necessarily restricted. If that subset does not depend on $\tilde{r}$, however, then the new protocol is restricted. It is not hard to verify that the subsets used (in the proof) never depend on $\tilde{r}$.

Theorem 1.1 is thus correct even if we allow only restricted protocols. Theorem 7.2 follows. □

Acknowledgments. I would like to thank Uri Feige and Moni Naor for presenting the problem to me, and Uri Feige, Peter Hajnal, Joe Kilian, Dieter Van Melkebeek, Mario Szegedy, Endre Szemeredi, Gabor Tardos, Oleg Verbitsky, and Avi Wigderson for helpful and encouraging discussions on the subject and on the content of this paper. I would like to thank the two referees for many helpful comments. This research was carried out while the author was a postdoc at Princeton University and DIMACS.

REFERENCES

[1] N. Alon, *Probabilistic methods in extremal finite set theory*, in Proc. Conference on Extremal Problems for Finite Sets, Bolyai Soc. Math. Stud. 3, János Bolyai Society, Budapest, Hungary, 1994, pp. 39–57.

[2] S. Arora, *Proof Verification and Hardness of Approximation Problems*, Ph.D. dissertation, University of California, Berkeley, CA, http://www.cs.princeton.edu/∼arora, 1994.

[3] S. Arora and C. Lund, *Hardness of approximations*, in Approximation Algorithms for NP-Hard Problems, D. Hochbaum, ed., PWS Publishing, Boston, MA, 1996; also available online from http://www.cs.princeton.edu/∼arora.

[4] S. Arora and S. Safra, *Probabilistic checking of proofs: A new characterization of NP*, in Proc. FOCS 1992, IEEE Computer Society Press, Los Alamitos, CA, pp. 2–13.

[5] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy, *Proof verification and intractability of approximation problems*, in Proc. FOCS 1992, IEEE Computer Society Press, Los Alamitos, CA, pp. 14–23.

[6] M. Bellare, *Interactive proofs and approximation*, in Proc. Israel Symposium on Theory of Computing and Systems, IEEE Computer Society Press, Los Alamitos, CA, 1993, pp. 266–274.

[7] L. Babai, L. Fortnow, and C. Lund, *Non-deterministic exponential time has two-prover interactive protocols*, in Proc. FOCS 1990, IEEE Computer Society Press, Los Alamitos, CA, pp. 16–25.

[8] M. Bellare, S. Goldwasser, C. Lund, and A. Russell, *Efficient probabilistic checkable proofs and applications to approximation*, in Proc. STOC 1993, ACM, New York, pp. 294–304.

[9] M. Bellare, O. Goldreich, and M. Sudan, *Free bits, PCPs, and nonapproximability—towards tight results*, SIAM J. Comput., 27 (1998), pp. 804–915.

[10] M. Bellare and P. Rogaway, *The complexity of approximating a nonlinear program*, Math. Programming, 69 (1995), pp. 429–441.

[11] M. Bellare and M. Sudan, *Improved non-approximability results*, in Proc. STOC 1994, ACM, New York, pp. 184–193.

[12] M. Ben-or, S. Goldwasser, J. Kilian, and A. Wigderson, *Multi prover interactive proofs: How to remove intractability*, in Proc. STOC 1988, ACM, New York, pp. 113–131.

[13] M. Ben-or, S. Goldwasser, J. Kilian, and A. Wigderson, *Efficient identification schemes using two prover interactive proofs*, in Proc. Crypto 1989, Springer-Verlag, New York, 1990, pp. 498–506.

[14] J. Cai, A. Condon, and R. Lipton, *On bounded round multi-prover interactive proof systems*, in Proc. Structure in Complexity Theory, 1990, IEEE Computer Society Press, Los Alamitos, CA, pp. 45–54.

[15] J. Cai, A. Condon, and R. Lipton, *Playing games of incomplete information*, Theoret. Comput. Sci., 103 (1992), pp. 25–38.

[16] J. Cai, A. Condon, and R. Lipton, *PSPACE is provable by two provers in one round*, J. Comput. System Sci., 48 (1994), pp. 183–193.

[17] I. Csiszar and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, New York, London, 1981.

[18] C. Dwork, U. Feige, J. Kilian, M. Naor, and S. Safra, *Low communication, 2-prover zero-knowledge proofs for NP*, in Proc. Crypto 1992, Springer-Verlag, Berlin, pp. 217–229.

[19] U. Feige, *On the success probability of the two provers in one round proof systems*, in Proc. Structures 1991, pp. 116–123.

[20] U. Feige, *Error Reduction by Parallel Repetition: The State of the Art*, Technical report CS95-32, Weizmann Institute of Science, Rehovot, Israel.

[21] L. Fortnow, *Complexity-Theoretic Aspects of Interactive Proof Systems*, Ph.D. thesis, Report MIT/LCS/TR-447, MIT, Cambridge, MA, 1989.

[22] U. Feige, S. Goldwasser, L. Lovasz, M. Safra, and M. Szegedy, *Approximating clique is almost NP-complete*, in Proc. FOCS 1991, IEEE Computer Society Press, Los Alamitos, CA, pp. 2–12.

[23] U. Feige and J. Kilian, *Two prover protocols: Low error at affordable rates*, in Proc. STOC 1994, ACM, New York, pp. 172–183.

[24] U. Feige and J. Kilian, *Impossibility results for recycling random bits in two prover proof systems*, in Proc. STOC 1995, ACM, New York, pp. 457–468.

[25] U. Feige and L. Lovasz, *Two-prover one-round proof systems, their power and their problems*, in Proc. STOC 1992, ACM, New York, pp. 733–744.

[26] L. Fortnow, J. Rompel, and M. Sipser, *On the power of multi-prover interactive protocols*, in Proc. Structures 1988, pp. 156–161.

[27] U. Feige and O. Verbitsky, *Error reduction by parallel repetition: A negative result*, in Proc. 11th Annual IEEE Conference on Computational Complexity, 1996, IEEE Computer Society Press, Los Alamitos, CA, pp. 70–76.

[28] L. Fortnow, J. Rompel, and M. Sipser, *Errata for "On the power of multi-prover interactive protocols,"* in Proc. Structure in Complexity Theory 1990, IEEE Computer Society Press, Los Alamitos, CA, pp. 318–319.

[29] R. M. Gray, *Entropy and Information Theory*, Springer-Verlag, New York, 1990.

[30] J. Håstad, *Testing of the long code and hardness for clique*, in Proc. STOC 1996, ACM, New York, pp. 11–19.

[31] J. Håstad, *Clique is Hard to Approximate Within* $n^{1-\epsilon}$, in Proc. FOCS 1996, IEEE Computer Society Press, Los Alamitos, CA, pp. 627–636.

[32] J. Håstad, *Some optimal inapproximability results*, in Proc. STOC 1997, ACM, New York, pp. 1–10.

[33] J. Kilian, *Strong Separation Models of Multi Prover Interactive Proofs*, DIMACS Workshop on Cryptography, October 1990.

[34] E. Kushilevitz and N. Nisan, *Communication Complexity*, Cambridge University Press, London, Cambridge, 1997.

[35] B. Kalyanasundaram and G. Schnitger, *The probabilistic communication complexity of set intersection*, in Proc. Structure in Complexity Theory 1987, IEEE Computer Society Press, Los Alamitos, CA, pp. 41–49.

[36] D. Lapidot and A. Shamir, *A one-round, two-rover, zero-knowledge protocol for NP*, in Combinatorica, 15 (1995), pp. 203–214.

[37] D. Lapidot and A. Shamir, *Fully parallelized multi prover protocols for NEXP-time*, in Proc. FOCS 1991, IEEE Computer Society Press, Los Alamitos, CA, pp. 13–18.

[38] C. Lund and M. Yannakakis, *On the hardness of approximating minimization problems*, in Proc. STOC 1993, ACM, New York, pp. 286–293.

[39] D. Peleg, *On the maximum density of* 0-1 *Matrices with no forbidden rectangles*, Discrete Math., 140 (1995), pp. 269–274.

[40] R. Raz, *A parallel repetition theorem*, in Proc. STOC 1995, ACM, New York, pp. 447–556.

[41] A. A. Razborov, *On the distributional complexity of disjointness*, Theoret. Comput. Sci., 106 (1992), pp. 385–390.

[42] I. Parnafes, R. Raz, and A. Wigderson, *Direct product results and the GCD problem in old and new communication models*, in Proc. STOC 1997, ACM, New York, pp. 363–372.

[43] G. Tardos, *Multi-prover encoding schemes, and three-prover proof systems*, J. Comput. System Sci., 53 (1996), pp. 251–260.

[44] O. Verbitsky, *Towards the parallel repetition conjecture*, Theoret. Comput. Sci., 157 (1996), pp. 277–282.

[45] O. Verbitsky, *The Parallel Repetition Conjecture for Trees is True*, manuscript, 1994.