Rounding Parallel Repetitions of Unique Games

Boaz Barak[*]
Princeton University

Moritz Hardt[†]
Princeton University

Ishay Haviv[‡]
Tel Aviv University

Anup Rao[§]
Institute for Advanced Study

Oded Regev[¶]
Tel Aviv University

David Steurer[‖]
Princeton University

Abstract

We show a connection between the semidefinite relaxation of unique games and their behavior under parallel repetition. Specifically, denoting by $\mathsf{val}(G)$ *the value of a two-prover unique game* G*, and by* $\mathsf{sdpval}(G)$ *the value of a natural semidefinite program to approximate* $\mathsf{val}(G)$*, we prove that for every* $\ell \in \mathbb{N}$*, if* $\mathsf{sdpval}(G) \geqslant 1 - \delta$*, then* $\mathsf{val}(G^\ell) \geqslant 1 - \sqrt{s\ell\delta}$ *. Here,* G^ℓ *denotes the* ℓ*-fold parallel repetition of* G*, and* $s = O(\log(k/\delta))$*, where* k *denotes the alphabet size of the game. For the special case where* G *is an XOR game (i.e.,* $k = 2$*), we obtain the same bound but with* s *as an absolute constant. Our bounds on* s *are optimal up to a factor of* $O(\log(1/\delta))$*.*

For games with a significant gap between the quantities $\mathsf{val}(G)$ *and* $\mathsf{sdpval}(G)$*, our result implies that* $\mathsf{val}(G^\ell)$ *may be much larger than* $\mathsf{val}(G)^\ell$*, giving a counterexample to the strong parallel repetition conjecture. In a recent breakthrough, Raz (FOCS '08) has shown such an example using the max-cut game on odd cycles. Our results are based on a generalization of his techniques.*

1. Introduction

In a *two-prover game*, a referee interacts with two provers, whose joint goal is to maximize the probability that the referee outputs "accept". The provers may decide in advance on an arbitrary strategy, and they may use shared randomness, but they cannot communicate with one another during the interaction, which proceeds as follows:

1. The referee samples a pair of queries (u, v) from a distribution $\mathcal{G}$ specified by the game.

2. The referee sends u to the first prover, and obtains an answer i, where $i \in [k]$ for some integer k that is called the *alphabet size* of the game.

3. The referee sends v to the second prover and obtains an answer $j \in [k]$.

4. The referee applies a predicate specified by the game to (u, v, i, j) and decides accordingly whether to accept or to reject.

The game is called *unique* if the predicate consists of checking whether $j = \pi_{uv}(i)$ where π_{uv} is a permutation of $[k]$. A unique game with alphabet size 2 is called an *XOR game*. The *value* of the game G, denoted by $\mathsf{val}(G)$, is the maximum probability of success that the provers can achieve (the provers succeed if the referee accepts their answers).

Two-prover games have turned out to be useful in several contexts, including hardness of approximation and quantum mechanics (see [12, 3, 8]). In many of these applications, it is important to understand how the value of the game decreases under *parallel repetition*. For $\ell \in \mathbb{N}$, the *ℓ-fold parallel repetition of G*, denoted by G^ℓ, is the game in which the referee samples ℓ pairs of questions $(u_1, v_1), \dots, (u_\ell, v_\ell)$ independently from $\mathcal{G}$, sending $(u_1, \dots, u_\ell)$ to the first prover and $(v_1, \dots, v_\ell)$ to the second prover. The provers then respond with $(i_1, \dots, i_\ell)$ and $(j_1, \dots, j_\ell)$, respectively. The referee accepts if and only if

[*] Department of Computer Science, Princeton University, boaz@cs.princeton.edu. Supported by NSF grants CNS-0627526 and CCF-0426582, US-Israel BSF grant 2004288 and Packard and Sloan fellowships.

[†] Department of Computer Science, Princeton University, mhardt@cs.princeton.edu.

[‡] School of Computer Science, Tel Aviv University, Tel Aviv, Israel. Supported by the Binational Science Foundation, and by the Israel Science Foundation.

[§] School of Mathematics, Institute for Advanced Study, arao@ias.edu. Supported in part by NSF Grant No. CCR-0324906.

[¶] School of Computer Science, Tel Aviv University, Tel Aviv, Israel. Supported by the Binational Science Foundation, by the Israel Science Foundation, by the European Commission under the Integrated Project QAP funded by the IST directorate as Contract Number 015848, and by a European Research Council (ERC) Starting Grant.

[‖] Department of Computer Science, Princeton University. Supported by NSF grants MSPA-MCS 0528414, and ITR 0205594.

each of the tuples $(u_1, v_1, i_1, j_1), \ldots, (u_\ell, v_\ell, i_\ell, j_\ell)$ is acceptable.

Clearly the two provers for G^ℓ can run the optimal strategy for the one-shot game in each instance independently, and thus $\mathsf{val}(G^\ell) \geqslant \mathsf{val}(G)^\ell$. Understanding the true value of $\mathsf{val}(G^\ell)$ is a fundamental question. The first dramatic progress was made by Raz [24] who proved that for every game G of alphabet size k with $\mathsf{val}(G) = 1 - \varepsilon$,

$$\mathsf{val}(G^\ell) \leqslant (1 - \varepsilon^c)^{\Omega(\ell/s)}, \tag{1}$$

where here $c = 32$ and $s = \log k$. Raz's argument was subsequently simplified by Holenstein [13], who also improved the bound (1) to get $c = 3$. This was followed by a work of Rao [23] showing that in the case of *projection games* (which are a generalization of unique games), the bound can be improved to $c = 2$ and $s = 1$.

It remained open whether the bound (1) could be further improved to yield $c = 1$, and the conjecture that this holds is known as the *strong parallel repetition* conjecture. As shown by Feige et al. [9], this conjecture would have had interesting applications to hardness of approximation results; moreover some of these applications would hold even if the conjecture was only true for the special case of unique games.

In a recent breakthrough, Raz [25] disproved the conjecture by giving an example of a game G of value $1 - \varepsilon$ for which $\mathsf{val}(G^\ell) \geqslant 1 - O(\sqrt{\ell}\,\varepsilon)$, which can be seen to imply that for $\ell \geqslant 1/\varepsilon^2$, $\mathsf{val}(G^\ell) \geqslant (1 - \varepsilon^2)^{O(\ell)}$. In addition, Raz's game is an XOR game (and a very simple one at that—the max-cut game on an odd cycle, see Section 2), and hence Raz disproved the strong parallel repetition conjecture even for this case. In this paper, we generalize Raz's results and techniques to show that a wide class of unique games yields such a counterexample. In fact, we determine asymptotically the value of the repeated version of *any* unique game, up to logarithmic factors in the exponent.

Semidefinite relaxation of games. While computing (or even approximating) the value of a two-prover game is **NP**-hard in general, Feige and Lovász [10] showed a non-trivial upper bound on this value by an efficiently computable parameter using semidefinite programming (SDP). For every game G, the problem of computing $\mathsf{val}(G)$ (i.e., maximizing the acceptance probability) can be formalized as a quadratic programming problem, and [10] studied a natural semidefinite relaxation of this program (see Figure 1).[1] Letting $\mathsf{sdpval}(G)$ be the optimum value of this program, it always holds that $\mathsf{sdpval}(G) \geqslant \mathsf{val}(G)$. Using duality it can be shown $\mathsf{sdpval}(G^\ell) = \mathsf{sdpval}(G)^\ell$ [10] (see also [21]). Hence, games with $\mathsf{val}(G) = \mathsf{sdpval}(G)$ have perfect parallel repetition, $\mathsf{val}(G^\ell) = \mathsf{val}(G)^\ell$. However, there are known examples with a large *gap* between $\mathsf{val}(G)$ and $\mathsf{sdpval}(G)$ and for such games the value of G^ℓ was not known. Our results imply that as ℓ grows, the value of G^ℓ tends not to $\mathsf{val}(G)^\ell$ but rather to $\mathsf{sdpval}(G)^\ell$. This has strong negative consequences for [9]'s intended application of using strong parallel repetition to show hardness of approximation, because for such applications the interesting games are those where $\mathsf{val}(G)$ is hard to approximate and hence is far from $\mathsf{sdpval}(G)$. We show that in all these cases strong parallel repetition fails to hold. (In fact, our results combined with [10] imply that the unique games where strong parallel repetition fails are exactly those with a large gap between $\mathsf{val}(G)$ and $\mathsf{sdpval}(G)$.)

1.1. Our Results

Our main result is the following:

Theorem 1.1. *For every $\ell \in \mathbb{N}$ and every unique game G with $\mathsf{sdpval}(G) \geqslant 1 - \delta$ and alphabet size k,*

$$\mathsf{val}(G^\ell) \geqslant 1 - O(\sqrt{\ell\delta \log(k/\delta)})\,.$$

For the special case of XOR games, we can prove a stronger bound that avoids the logarithmic factor.

Theorem 1.2. *For every $\ell \in \mathbb{N}$ and every XOR game G with $\mathsf{sdpval}(G) \geqslant 1 - \delta$, we have $\mathsf{val}(G^\ell) \geqslant 1 - 4\sqrt{\ell\delta}$.*

When the number of repetitions ℓ tends to infinity, we obtain the following stronger bound. Here we use $\overline{\mathsf{val}}(G)$ to denote the *asymptotic value of G*, defined as $\lim_{\ell\to\infty} (\mathsf{val}(G^\ell))^{1/\ell}$.

Theorem 1.3. *For every XOR game G with $\mathsf{sdpval}(G) \geqslant 1 - \delta$, we have $\overline{\mathsf{val}}(G) \geqslant 1 - 2\delta$.*

The proofs of Theorem 1.1 and Theorem 1.2 are presented in Section 4.4. For details about Theorem 1.3, see the full version of the paper.

For every ℓ and every m that divides ℓ, two provers playing the game G^ℓ can always achieve acceptance probability at least $\mathsf{val}(G^m)^{\ell/m}$ by using the optimal strategy on each block of m repetitions independently. Combining this observation with Feige and Lovász's [10] result that $\mathsf{val}(G^\ell) \leqslant \mathsf{sdpval}(G)^\ell$, we obtain the following corollary to Theorems 1.1 and 1.2:

Corollary 1.4. *For every unique two-prover game G with $\mathsf{sdpval}(G) = 1 - \delta$ and for every $\ell > 1/\delta$,*

$$\mathsf{val}(G^\ell) = (1 - \delta)^{\tilde{\Theta}(\ell)}, \tag{2}$$

where the $\tilde{\Theta}$ notation hides factors logarithmic in $1/\delta$ and the alphabet size of G. Moreover, if G is an XOR game then we can replace the right-hand side of (2) with $(1 - \delta)^{\Theta(\ell)}$.

[1] The SDP of [10] had some additional constraints above those of the SDP of Figure 1, but for our purposes they are still essentially equivalent. For details see the full version of the paper.

The dependence on $\log k$ in our bound is inherent, as can be seen by combining the Khot–Vishnoi [17] integrality gap example with Rao's parallel repetition theorem for unique games [23]: the results in [17] and [16] allow to construct a unique game G for which $\mathsf{sdpval}(G) = 1 - \delta$ but $\mathsf{val}(G) \leqslant 1 - O(\sqrt{\delta \log k})$ (for details of this construction, see the full version of the paper); Rao's parallel repetition theorem [23] implies that for this game $\mathsf{val}(G^\ell) \leqslant (1 - \delta \log k)^{\Omega(\ell)} = (1 - \delta)^{\Omega(\ell \log k)}$. Hence, the dependence on $\log k$ in Theorem 1.1 is optimal. In contrast, we conjecture that the dependence on $\log(1/\delta)$ in Theorem 1.1 is not inherent and is an artifact of our specific construction. Indeed, for the case of unique games with linear constraints we can remove the dependence on $\log(1/\delta)$, albeit at the cost of a worse dependence on k (for details, see Section 4.3 and the full version of the paper).

Previous rounding algorithms. It is instructive to compare our results with the best known rounding algorithm for unique games in this parameter regime, namely the CMM algorithm of Charikar, Makarychev and Makarychev [5]. That algorithm shows that if $\mathsf{sdpval}(G) \geqslant 1 - \delta$ then $\mathsf{val}(G) \geqslant 1 - O(\sqrt{\delta \log k})$, where k is G's alphabet size. Since the alphabet size of G^ℓ is k^ℓ and $\mathsf{sdpval}(G^\ell) = \mathsf{sdpval}(G)^\ell \sim 1 - \ell\delta$ in the range where this bound is nontrivial, the CMM algorithm on G^ℓ simply shows that

$$\mathsf{val}(G^\ell) \geqslant 1 - O(\sqrt{\ell\delta\ell \log k}) = 1 - O(\ell \sqrt{\delta \log k}) . \qquad (3)$$

Hence the CMM algorithm does not give any better guarantee on a repeated game than can be given by applying the algorithm on each coordinate separately (which is not surprising, as otherwise they would have already refuted the strong parallel repetition conjecture). In contrast, our Theorem 1.1 gives a bound of $1 - O(\sqrt{\ell\delta(\log k + \log(1/\delta)})$ which could be significantly better (i.e., closer to 1) than the right-hand side of (3) if $\log(1/\delta) \ll \ell$ (typically we think of $\ell \sim 1/\delta$ in which case $\log(1/\delta) \sim \log \ell$).

Games with entanglement. There are several known examples of games (often unique) in which provers who share quantum entanglement can achieve success probability higher than that achievable by provers without entanglement. Such games are used in the context of quantum information theory as experiments that validate some of the predictions of quantum mechanics. Kempe, Regev, and Toner [14] recently showed that the success probability achievable by entangled provers in unique games can be closely approximated by an SDP. Their proof involves a rounding strategy that produces strategies for provers with entanglement. Since it is known that the value achievable by entangled provers is always upper bounded by $\mathsf{sdpval}(G)$, our results show that as the number of repetitions ℓ grows, the ℓ^{th} root of the success probability that classical provers can achieve approaches (up to logarithmic factors) the ℓ^{th} root of the success probability that quantum provers can achieve. Thus to a certain extent the gap between quantum and classical provers in unique games can "shrink" with the number of repetitions.

Bounds for particular games. Our methods can be used to derive improved lower bounds on the amortized value of particular games. An especially interesting example is the XOR game known as the CHSH game, introduced by Clauser et al. [8] in 1969. In this game the referee sends a uniform and independent bit to each prover, and each prover responds with a value from $\{-1, 1\}$. The constraint is an inequality constraint if and only if both questions are 1. By always answering 1, the provers can win with probability $3/4$, and it is easy to see that this is the best possible strategy. It is well-known that the SDP value of this game equals $1/2 + 1/\sqrt{8} \approx 0.8535$. It is somewhat surprising that the asymptotic value $\overline{\mathsf{val}}(\text{CHSH})$ is not known. Aaronson pointed out that this value is at least $\sqrt{10}/4 \approx 0.7906$ by considering CHSH^2 (see [1, 2]). In the full version of this paper, we show that this asymptotic value is at least $\cos(\pi/5) = 1/4(1 + \sqrt{5}) \approx 0.809$.

2. Techniques

Our techniques are a natural generalization of Raz's work [25], and so it is instructive to start with a high level overview of his approach. Raz's counterexample used the following simple XOR game. In the *max-cut* game on a graph G the referee selects a random edge $\{u, v\}$ of G, sets $x = u$, then with probability $1/2$ sets $y = u$ and with probability $1/2$ sets $y = v$. The referee sends x, y to the provers and receives two bits a, b from them respectively. If $x = y$ then it accepts if $a = b$, and if $x \neq y$ then it accepts if $a \neq b$. The game is called the *max-cut* game since (as can be easily seen) if the maximum cut in G cuts a $1 - \varepsilon$ fraction of the edges, then the value of the game is $1 - \varepsilon/2$. In particular, if G is the n vertex cycle for some odd n, then the value of the corresponding game (which we denote also by G) is $1 - 1/(2n)$. Raz achieved his counterexample by showing that in this case, for every ℓ the value of G^ℓ is at least $1 - O(\sqrt{\ell}/n)$.

Interestingly, a central tool used by Raz is a *correlated sampling* lemma that Holenstein [13] used to *prove* the parallel repetition theorem (for general games and with $c = 3$). We will give the formal statement of the correlated sampling lemma below (Lemma 4.1) but roughly speaking, it says that if the two provers are given a pair of distributions D, D' with statistical distance at most ε, then even without communicating, each prover can sample a random element according to his distribution such that with probability $1-2\varepsilon$ both provers output the *same* element.

Raz used the correlated sampling lemma in the following way. He defined for every vertex u in the odd cycle a distribution D_u over cuts in the graph, such that the probability that a cut selected from D_u does not cut one of the two edges touching u is very small (i.e., $O(1/n^2)$). The provers' strategy on input u is to sample a cut (S, T) according to D_u and output either 1 or 0 according to whether or not $u \in S$. Now if it happened to be the case that the cut sampled by the first prover on input u is the same cut as the one sampled by the second prover on input v, it would mean that if $u = v$ then they answer the same value and if (u, v) is an edge then with high probability they answer a different value. Using the correlated sampling lemma, one can ensure that as long as the statistical distance $\Delta(D_u, D_v)$ of D_u and D_v is at most ε for every neighboring u, v in the graph, Raz's strategy will achieve success probability $1 - O(\varepsilon)$ (we are neglecting here the small probability that the cut misses the edge (u, v)). It can then be shown that the question of the value of the game under parallel repetition reduces to the question of the distance of many independent samples from these distributions.

Hellinger distance and independent samples. For two distributions D_1, D_2 of statistical distance ε, the statistical distance of D_1^ℓ and D_2^ℓ (where this denotes concatenating ℓ independent samples) depends quite a bit on the shape of the underlying distributions D_1 and D_2. For example, if D_1 is constantly 1 and D_2 is the biased coin with $\Pr[D_2 = 1] = 1 - \varepsilon$, then $\Delta(D_1, D_2) = \varepsilon$ and $\Delta(D_1^\ell, D_2^\ell) = 1 - (1 - \varepsilon)^\ell \approx \ell\varepsilon$ for small ℓ. On the other hand, if D_1 and D_2 are coins such that $\Pr[D_1 = 1] = 1/2 + \varepsilon/2$ and $\Pr[D_2 = 1] = 1/2 - \varepsilon/2$ then $\Delta(D_1, D_2)$ is also equal to ε, but $\Delta(D_1^\ell, D_2^\ell) = O(\sqrt{\ell}\,\varepsilon)$.[2] Raz uses in his paper a specific example of distributions D_1 and D_2 (that could be used for his provers' strategy) such that $\Delta(D_1, D_2) = \Theta(1/n)$ but $\Delta(D_1^\ell, D_2^\ell) = O(\sqrt{\ell}/n)$. In this paper, we note that the behavior of the statistical distance of product of distributions is determined by a different distance measure called the *Hellinger distance* (see Section 3). This distance measure has a geometric interpretation, which we use to relate it to the vector solution of the semidefinite program.

More concretely, we use Raz's approach in the context of *rounding* algorithms for unique games. Such algorithms transform a solution to the semidefinite program into a valid strategy for the original game. The rounding algorithms we use involve the provers selecting a random high dimensional vector for every input they receive from the verifier.[3] Specifically, we will define for every input u a distribution D_u on vectors such that if the two provers on input of u and v sample from the distributions D_u and D_v using the correlated sampling lemma, then the result is very likely to satisfy the predicate of the referee. The success probability of the provers depends crucially on the statistical distance between the distributions D_u and D_v. But in order to bound the statistical distance we will actually derive a bound on the Hellinger distance between the two distributions. This has the advantage that the bound carries over nicely to the *parallel repeated* game using simple properties of the Hellinger distance. We then use the quadratic relation between the Hellinger distance and the statistical distance to obtain a two-prover strategy for the repeated game. Finally, we show how solutions to the semidefinite programming relaxation give rise to distributions with small Hellinger distance and hence a good two-prover strategy for the repeated game. Of course this high level description is glossing over some very important details, (including the choice of distributions and rounding algorithms) and these are covered in the following sections.

3. Preliminaries

We use boldface to denote vectors. We will often use collections of vectors that are indexed by elements of some set $\mathcal{V}$. In this case, we write $\boldsymbol{u}$ for the vector indexed by the element $u \in \mathcal{V}$.

Statistical distance. Let X and Y be two probability distributions over a domain Ω (e.g., $[0, 1]$ or $\mathbb{R}^d$). Assume X and Y have density functions with respect to some measure μ (e.g., the Lebesgue measure), and let f and g denote these density functions.[4] We define their *statistical distance* (also known as *total variation distance*) by

$$\Delta(X, Y) \overset{\text{def}}{=} \frac{1}{2} \int_\Omega |f - g| \, \mathrm{d}\mu \, .$$

Notice that for any X and Y, $\Delta(X, Y) \in [0, 1]$.

Hellinger distance. For X and Y as above, one defines their *Hellinger distance* $\mathrm{H}(X, Y)$ as the square root of

$$\mathrm{H}^2(X, Y) \overset{\text{def}}{=} \frac{1}{2} \int_\Omega \left(\sqrt{f} - \sqrt{g} \right)^2 \mathrm{d}\mu = 1 - \int_\Omega \sqrt{fg} \, \mathrm{d}\mu \, .$$

[2] The best distinguisher between D_1^ℓ and D_2^ℓ will simply see whether the sum of the ℓ samples is larger than $\ell/2$. The difference in expectation between these two cases is $\varepsilon\ell$, which is equal to $\Theta(\varepsilon\ell/\sqrt{\ell}) = \Theta(\varepsilon\sqrt{\ell})$ standard deviations.

[3] We note that this is in contrast to standard rounding algorithms for semidefinite programs that typically select some global vectors to use in all cases; our case is different since the distribution of vectors we choose depends on the input query (i.e., vertex in the case of games on graphs).

[4] In more precise terms, we require X and Y to be absolutely continuous with respect to μ, and we let f and g be their Radon-Nikodym derivatives with respect to μ.

In other words, the Hellinger distance is the Euclidean distance between the unit vectors obtained from the density functions by taking the square root. We will mostly work with the square of the Hellinger distance, $\mathrm{H}^2(X, Y)$. Notice that for any X and Y, $\mathrm{H}^2(X, Y) \in [0, 1]$.

We will use the following known facts about the Hellinger distance. The first lemma relates the Hellinger distance to the total variation distance. In the second lemma, we see how the Hellinger distance of product distributions behaves.

Lemma 3.1 ([20, 22])**.** *For any two distributions X and Y,*

$$\mathrm{H}^2(X, Y) \leqslant \Delta(X, Y) \leqslant \sqrt{\mathrm{H}^2(X, Y)(2 - \mathrm{H}^2(X, Y))} \leqslant \sqrt{2}\,\mathrm{H}(X, Y)\,.$$

Lemma 3.2. *Let $\{X_1, \ldots, X_\ell\}$ and $\{Y_1, \ldots, Y_\ell\}$ be two families of distributions. Then,*

$$\mathrm{H}^2(X_1 \otimes \cdots \otimes X_\ell, Y_1 \otimes \cdots \otimes Y_\ell) = 1 - \prod_{i=1}^{\ell}(1 - \mathrm{H}^2(X_i, Y_i)) \leqslant \sum_{i=1}^{\ell} \mathrm{H}^2(X_i, Y_i)\,.$$

Here, $X_1 \otimes \cdots \otimes X_\ell$ denotes the product of the distributions $X_1, \ldots, X_\ell$, i.e., the distribution obtained by taking independent samples of $X_1, \ldots, X_\ell$.

Note that as a corollary of these two lemmas we obtain that for any two distributions D_1 and D_2, $\Delta(D_1^\ell, D_2^\ell) = O(\sqrt{\ell}\,\mathrm{H}(D_1, D_2))$.

The Hellinger distance defines a metric on distributions.

Lemma 3.3. *For any three distributions X, Y, and Z,*

$$\mathrm{H}(X, Y) \leqslant \mathrm{H}(X, Z) + \mathrm{H}(Z, Y)\,.$$

Finally, we state a useful lemma about the Hellinger distance between convex combinations.

Lemma 3.4. *Let X be a convex combination of the distributions $\{X_1, \ldots, X_\ell\}$ with coefficients α_i and let Y be a convex combination of $\{Y_1, \ldots, Y_\ell\}$ with coefficients β_i. Then,*

$$\mathrm{H}^2(X, Y) \leqslant \textstyle\sum_i \sqrt{\alpha_i \beta_i}\,\mathrm{H}^2(X_i, Y_i) + \mathrm{H}^2(\alpha, \beta).$$

We omit the (simple) proofs of these lemmas from this extended abstract (see the survey [11] and the references therein).

3.1. Unique Games and Semidefinite Relaxation

We represent a *unique game* G as a distribution over triples (u, v, π) where u and v are *queries* and π is a permutation of the *alphabet* $[k]$ of G. This representation differs slightly from the one used in earlier work, but it turns out to be very convenient, especially when dealing with parallel repetition. We say that a game is a two-prover game if the supports of the first and second component are disjoint. Our results all hold for general (not necessary two-prover) unique games, but note that essentially all known upper bounds on the value of repeated unique games are known to hold only in the two-prover case [10, 24, 13, 23].

$$\begin{array}{lll} \textbf{Maximize} & \mathop{\mathsf{E}}_{(u,v,\pi)\sim G} \sum_{i\in[k]} \langle \boldsymbol{u}_i, \boldsymbol{v}_{\pi(i)} \rangle & (4) \\ \textbf{Subject to} & \sum_{i\in[k]} \|\boldsymbol{u}_i\|_2^2 = 1 \qquad (u \in \mathcal{V}) & (5) \\ & \langle \boldsymbol{u}_i, \boldsymbol{u}_j \rangle = 0 \qquad (u \in \mathcal{V},\ i \neq j \in [k]) & (6) \end{array}$$

Figure 1. A semidefinite programming formulation of a unique game G. The variables are vectors $\boldsymbol{u}_i$ for every query u and every $i \in [k]$. Notice that the objective function can be equivalently written as $1 - \frac{1}{2}\mathsf{E}_{(u,v,\pi)\sim G} \sum_{i\in[k]} \|\boldsymbol{u}_i - \boldsymbol{v}_{\pi(i)}\|_2^2$.

We denote by $\mathcal{V}$ the set of possible queries, i.e., the support of the first and second components of the distribution G. A *solution* (also called *strategy*) for the game G is a collection $\{l_u\}_{u\in\mathcal{V}}$ of *labels* in $[k]$. The *value* of such a solution is the probability

$$\Pr_{(u,v,\pi)\sim G}[\pi(l_u) = l_v]\,. \qquad (7)$$

The maximum of (7) over all possible solutions is denoted by $\mathsf{val}(G)$.

The semidefinite program given in Figure 1 is a natural relaxation of the value of a game. Let $\mathsf{sdpval}(G)$ denote the optimum of this program. To see why $\mathsf{sdpval}(G) \geqslant \mathsf{val}(G)$ for any game G, notice that any solution $\{l_u\}_{u\in\mathcal{V}}$ of G can be converted into a feasible solution of the SDP by setting for each $u \in \mathcal{V}$ the vector $\boldsymbol{u}_{l_u}$ corresponding to the label l_u to some globally fixed unit vector, and all other $k - 1$ vectors to zero.

The *ℓ-fold repetition* of a game corresponds to the ℓ-fold product of the distribution G. We use the notations $\underline{u} = (u^{(1)}, \ldots, u^{(\ell)})$ and $\underline{\pi} = (\pi^{(1)}, \ldots, \pi^{(\ell)})$ to denote the queries and permutations of G^ℓ, respectively. If G is a two-prover game, then following an approach of Feige and Lovász [10] one can show that $\mathsf{sdpval}(G^\ell) = \mathsf{sdpval}(G)^\ell$ [21, 14].

4. Correlated Distributions, Repeated Unique Games, and Hellinger Distances

In this section we give our main “meta lemma” (Lemma 4.5) that allows to derive strategies for two-prover unique games from families of distributions with bounded Hellinger distance. Our results for XOR and general unique games will be derived by "plugging in" suitable distributions into this lemma. (Raz's result [25] can also be viewed in this framework.) The main tool we use is the *correlated sampling lemma*.

4.1. The Correlated Sampling Lemma

Consider two computationally unbounded provers that share a source of randomness Z but cannot communicate with each other. Assume we have some finite family of distributions $\{R_u\}_{u\in\mathcal{V}}$ over a domain Ω. The first prover is given an index $u \in \mathcal{V}$, the second prover is given $v \in \mathcal{V}$ and they want to sample an element $r_u(Z)$ and $r_v(Z)$ from distributions R_u and R_v, respectively. The next lemma shows that using shared randomness, the provers can correlate their samples such that provided R_u and R_v are statistically close, they end up with the same sample (i.e., $r_u(Z) = r_v(Z)$) with high probability.

Lemma 4.1. *Let $\{R_u\}_{u\in\mathcal{V}}$ be a family of distributions over some domain Ω. Then, there exists a family of functions $\{r_u : [0,1] \to \Omega\}_{u\in\mathcal{V}}$ such that if Z is a random variable uniformly distributed in $[0,1]$, then for every $u \in \mathcal{V}$, $r_u(Z)$ is distributed according to R_u, and for every $u, v \in \mathcal{V}$,*

$$\Pr[r_u(Z) = r_v(Z)] = \frac{1 - \Delta(R_u, R_v)}{1 + \Delta(R_u, R_v)} \geqslant 1 - 2\Delta(R_u, R_v).$$

In this paper we actually use a continuous version of the lemma, but it can be easily reduced to the discrete lemma by using a sufficiently fine discretization of the domain. (We omit the details in this extended abstract.)

The proof of the lemma uses a technique that has been used in several instances in computer science. Broder [4] used this technique for sketching sets, while the (discrete version of the) correlated sampling lemma was first proven by Kleinberg and Tardos [19] in the context of rounding algorithms for linear programs (see also [6, Sec 4.1]). It was rediscovered and used in the proof of the parallel repetition theorem by Holenstein [13].

The idea of the proof is simplest to describe in the case that every distribution R_u is uniform over some set S_u from a finite universe, and the sets S_u all have the same cardinality. In this case the provers can simply interpret the shared randomness as a random ordering of the universe and each prover outputs on input u the element $r_u(Z)$ that is the minimal element in S_u according to this order. Clearly, $r_u(Z)$ is distributed uniformly in S_u. On the other hand, $\Pr[r_u(Z) = r_v(Z)] = |S_u \cap S_v|/|S_u \cup S_v|$, which is equal to $(1 - \Delta(R_u, R_v))/(1 + \Delta(R_u, R_v))$. For an arbitrary distribution R, we can emulate the previous approach by duplicating every element r in the support of R a number of times that is proportional to $\Pr[R = r]$.

4.2. From Correlated Distributions to Strategies for Unique Games

Definition 4.2. A family $\{X_u\}_{u\in\mathcal{V}}$ of distributions of the form $X_u = (R_u, L_u)$ supported on $\Omega \times [k]$ is called a *distributional strategy* if for every $u \in \mathcal{V}$, L_u is a function of R_u, i.e., for every $u \in \mathcal{V}$ and every r in the support of R_u, there is exactly one $i \in [k]$ such that the pair (r, i) is in the support of the distribution X_u. Equivalently, we can say that for every $u \in \mathcal{V}$ and every $i \neq j \in [k]$, the supports of the conditional distributions $[R_u \mid L_u = i]$ and $[R_u \mid L_u = j]$ are disjoint.

We can think of the random variable R_u as a *random seed* that determines a *label* L_u for the query $u \in \mathcal{V}$.

If a unique game G has value at least $1 - \varepsilon$, it is easy to construct a distributional strategy $\{X_u\}_{u\in\mathcal{V}}$ such that

$$\mathop{\mathsf{E}}_{(u,v,\pi)\sim G} \Delta(\pi.X_u, X_v) \leqslant \varepsilon\,, \tag{8}$$

where $\pi.X_u$ denotes the distribution obtained from $X_u = (R_u, L_u)$ by applying π to the second component, that is,

$$\pi.X_u \overset{\text{def}}{=} (R_u, \pi(L_u)).$$

For instance, take $\Omega = \{1\}$ to be a singleton, and set each X_u to be constantly $(1, l_u)$ where $l_u \in [k]$ is the label given to u.

On the other hand, the next lemma shows that if a distributional strategy satisfies (8) for a game G, then G has value at least $1 - 2\varepsilon$.

Lemma 4.3. *Suppose $\{X_u\}_{u\in\mathcal{V}}$ is a distributional strategy for a unique game G. Then,*

$$\mathsf{val}(G) \geqslant 1 - 2 \mathop{\mathsf{E}}_{(u,v,\pi)\sim G} \Delta(\pi.X_u, X_v)\,. \tag{9}$$

The proof relies on the correlated sampling lemma (Lemma 4.1). A distributional strategy $\{X_u\}_{u\in\mathcal{V}}$ is rounded to a solution $\{l_u\}_{u\in\mathcal{V}}$ as follows: We apply Lemma 4.1 to obtain functions $r_u : [0,1] \to \Omega$. Then, we choose a random number Z uniformly from $[0,1]$. Now, for every $u \in \mathcal{V}$, we can uniquely determine a label l_u such that $(r_u(Z), l_u)$ is in the support of X_u. These labels $\{l_u\}_{u\in\mathcal{V}}$ form our solution. For the detailed proof, see Appendix A.1.

Remark 4.4. The above lemma is implicit in the analysis of an approximation algorithm for unique games by Chlamtac, Makarychev, and Makarychev [7]. Their algorithm obtains strategies for unique games from certain embeddings into L_1-space. It is easy to construct such an embedding from a distributional strategy. For details, see the full version of the paper.

Lemma 4.5. *Let G be a unique game. Suppose there exists a distributional strategy $\{X_u\}_{u\in\mathcal{V}}$ such that*

$$\mathop{\mathsf{E}}_{(u,v,\pi)\sim G} \mathrm{H}^2(\pi.X_u, X_v) \leqslant \delta\,. \tag{10}$$

Then, for every $\ell \in \mathbb{N}$, the ℓ-fold repetition of G has value

$$\mathsf{val}(G^{\ell}) \geqslant 1 - 2\sqrt{2\ell\delta}\,.$$

Proof. For $\underline{u} = (u^{(1)}, \ldots, u^{(\ell)}) \in \mathcal{V}^{\ell}$, let $X_{\underline{u}}$ denote the product distribution $X_{u^{(1)}} \otimes \cdots \otimes X_{u^{(\ell)}}$. Note that $\{X_{\underline{u}}\}_{\underline{u}\in\mathcal{V}^{\ell}}$ is a distributional strategy for the ℓ-fold repeated game G^{ℓ}. By Lemma 3.2 (subadditivity of H^2 for product distributions), the bound (10) implies that

$$\mathop{\mathsf{E}}_{(\underline{u},\underline{v},\underline{\pi})\sim G^{\ell}} \mathrm{H}^2(\underline{\pi}.X_{\underline{u}}, X_{\underline{v}})$$
$$\leqslant \mathop{\mathsf{E}}_{(\underline{u},\underline{v},\underline{\pi})\sim G^{\ell}} \mathrm{H}^2(\pi^{(1)}.X_{u^{(1)}}, X_{v^{(1)}}) + \cdots + \mathrm{H}^2(\pi^{(\ell)}.X_{u^{(\ell)}}, X_{v^{(\ell)}}) \leqslant \ell\delta\,.$$

Hence, by Lemma 3.1 and the concavity of the square root function,

$$\mathop{\mathsf{E}}_{(\underline{u},\underline{v},\underline{\pi})\sim G^{\ell}} \Delta(\underline{\pi}.X_{\underline{u}}, X_{\underline{v}}) \leqslant \sqrt{2\ell\delta}\,,$$

which implies by Lemma 4.3 that $\mathsf{val}(G^{\ell}) \geqslant 1 - 2\sqrt{2\ell\delta}$. □

It is crucial for our results that the above lemma is based on the square of the Hellinger distance, and not on the total variation distance, since the former can be quadratically smaller than the latter. Moreover, as we shall see next, the square of the Hellinger distance can be related to the objective function of the semidefinite relaxation.

4.3. From SDP Solutions to Correlated Distributions

In this section, we state three results that will allow us to relate the left-hand side of (10) to one minus the value of an optimal solution of the semidefinite program in Figure 1. We present proofs for the first two lemmata in Section 5. The proof for the third lemma is deferred to the full version of the paper.

Lemma 4.6. *Let $t \geqslant 1$ and let $\{\boldsymbol{u}_i\}_{u\in\mathcal{V}, i\in[k]}$ be a feasible solution of the SDP in Figure 1. Then, there exists a distributional strategy $\{X_u\}_{u\in\mathcal{V}}$ such that for every triple (u, v, π),*

$$\mathrm{H}^2(\pi.X_u, X_v) \leqslant O(t) \cdot \textstyle\sum_{i\in[k]} \frac{1}{2}\|\boldsymbol{u}_i - \boldsymbol{v}_{\pi(i)}\|^2 + k \cdot 2^{-t}\,. \tag{11}$$

For the case $k = 2$, a stronger upper bound holds.

Lemma 4.7. *Let $k = 2$ and let $\{\boldsymbol{u}_i\}_{u\in\mathcal{V}, i\in[k]}$ be a feasible solution of the SDP in Figure 1. Then, there exists a distributional strategy $\{X_u\}_{u\in V}$ such that for every triple (u, v, π),*

$$\mathrm{H}^2(\pi.X_u, X_v) \leqslant 2 \cdot \textstyle\sum_{i\in[k]} \frac{1}{2}\|\boldsymbol{u}_i - \boldsymbol{v}_{\pi(i)}\|^2\,.$$

Let Γ be some Abelian group of order k (e.g., $\mathbb{Z}_k$) whose elements are identified with $[k]$ in some fixed arbitrary way. We say that a permutation π on Γ is a Γ-*shift* if there exists an $s \in \Gamma$ such that for all $a \in \Gamma$, $\pi(a) = a + s$. Unique games that use only Γ-shifts are known as *linear games* or Γ-Max2Lin(k) *instances* (e.g., [14, 16]). For such games, the additive term $k \cdot 2^{-t}$ in (11) can be avoided at the cost of a worse multiplicative dependency on k.

Lemma 4.8. *Let $\{\boldsymbol{u}_i\}_{u\in\mathcal{V}, i\in[k]}$ be a feasible SDP solution and let Γ be as above. Then there exists a distributional strategy $\{X_u\}_{u\in\mathcal{V}}$ such that for every triple (u, v, π) with π a Γ-shift,*

$$\mathrm{H}^2(\pi.X_u, X_v) \leqslant c_k \cdot \textstyle\sum_{i\in[k]} \frac{1}{2}\|\boldsymbol{u}_i - \boldsymbol{v}_{\pi(i)}\|^2\,, \tag{12}$$

where c_k is a factor depending only on k.

Remark 4.9. In order to illustrate the basic idea of the construction of distributional strategies from SDP solutions, let us consider the following special case. Suppose $\{\boldsymbol{u}_i\}_{u\in\mathcal{V}, i\in[k]}$ is a feasible SDP solution such that every vector $\boldsymbol{u}_i \in \mathbb{R}^d$ has only *nonnegative* coordinates (with respect to the canonical basis of $\mathbb{R}^d$). Then, let $\{X_u\}_{u\in\mathcal{V}}$ be the distributional strategy such that

$$\Pr[X_u = (r, i)] = \boldsymbol{u}_i(r)^2\,, \tag{13}$$

for $r \in [d]$ and $i \in [k]$ where $\boldsymbol{u}_i(r)$ denotes the r^{th} coordinate of $\boldsymbol{u}_i$. The above equation specifies probability distributions over $[d] \times [k]$, because $\sum_{i\in[k], r\in[d]} \Pr[X_u = (r, i)] = \sum_{i\in[k]} \|\boldsymbol{u}_i\|^2 = 1$. The nonnegativity and orthogonality of $\boldsymbol{u}_i$ and $\boldsymbol{u}_j$ imply that the conditional distributions $[R_u \mid L_u = i]$ and $[R_u \mid L_u = j]$ have disjoint support. Hence, $\{X_u\}_{u\in\mathcal{V}}$ is indeed a distributional strategy. On the other hand, we have for every triple (u, v, π).

$$\mathrm{H}^2(\pi.X_u, X_v) = \sum_{i\in[k]} \tfrac{1}{2}\|\boldsymbol{u}_i - \boldsymbol{v}_{\pi(i)}\|^2\,. \tag{14}$$

If the vectors $\{\boldsymbol{u}_i\}_{u\in\mathcal{V}, i\in[k]}$ have negative entries, the right hand side of (14) is still an upper bound on $\mathrm{H}^2(\pi.X_u, X_v)$. However, the family of distributions $\{X_u\}_{u\in\mathcal{V}}$ constructed in (13) will in general fail to be a distributional strategy if the solution vectors have negative entries.

4.4. Putting it Together

Combining Lemma 4.3 and the Lemmata 4.6–4.8, we get the following theorem which implies Theorem 1.1 (rounding parallel repetitions of general unique games) and Theorem 1.2 (rounding parallel repetitions of XOR games).

Theorem 4.10. *For every $\ell \in \mathbb{N}$ and every unique game G with $\mathsf{sdpval}(G) \geqslant 1 - \delta$, we have $\mathsf{val}(G^{\ell}) \geqslant 1 - 2\sqrt{2s\ell\delta}$, where*

- *$s = O(\log(k/\delta))$ if G is a unique game with alphabet $[k]$,*
- *$s = 2$ if G is an XOR game (i.e., $k = 2$),*
- *$s = c_k$ if G is an instance of Γ-Max2Lin(k).*

Proof. Suppose G is a unique game on alphabet $[k]$ with $\mathsf{sdpval}(G) \geqslant 1-\delta$. Let $\{\boldsymbol{u}_i\}_{u\in\mathcal{V},i\in[k]}$ be an optimal SDP solution for G. Note that

$$\mathop{\mathsf{E}}_{(u,v,\pi)\sim G} \tfrac{1}{2}\|\boldsymbol{u}_i - \boldsymbol{v}_{\pi(i)}\|^2 = 1 - \mathop{\mathsf{E}}_{(u,v,\pi)\sim G} \langle \boldsymbol{u}_i, \boldsymbol{v}_{\pi(i)}\rangle \leqslant \delta\,.$$

We apply Lemma 4.6 for $t = \log(k/\delta)$ to obtain a distributional strategy that satisfies

$$\mathop{\mathsf{E}}_{(u,v,\pi)\sim G} \mathrm{H}^2(\pi.X_u, X_v) \leqslant O(t)\cdot\delta + k\cdot 2^{-t} = O(\delta\log(k/\delta))\,.$$

Now Lemma 4.5 implies that $\mathsf{val}(G^\ell) \geqslant 1 - O(\sqrt{\ell\delta\log(k/\delta)})$ for any $\ell \in \mathbb{N}$. The proof for the case that G is an XOR game or an instance of Γ-Max2Lin(k) is the same; the only change is that instead of Lemma 4.6 we apply Lemma 4.7 or Lemma 4.8 to obtain a distributional strategy. □

5. Constructions of Correlated Distributions from SDP Solutions

5.1. Proof of Lemma 4.7

Lemma 4.7 (Restated)**.** *Let $k = 2$ and let $\{\boldsymbol{u}_i\}_{u\in\mathcal{V},i\in[k]}$ be a feasible solution of the SDP in Figure 1. Then, there exists a distributional strategy $\{X_u\}_{u\in V}$ such that for every triple (u, v, π),*

$$\mathrm{H}^2(\pi.X_u, X_v) \leqslant 2\cdot \textstyle\sum_{i\in[k]} \tfrac{1}{2}\|\boldsymbol{u}_i - \boldsymbol{v}_{\pi(i)}\|^2\,.$$

Let $\{\boldsymbol{u}_i\}_{u\in\mathcal{V},i\in[k]}$ be a feasible SDP solution with $\boldsymbol{u}_i \in \mathbb{R}^d$. For each $u \in \mathcal{V}$, let $\boldsymbol{u}$ denote the unit vector $\boldsymbol{u}_1 - \boldsymbol{u}_2 \in \mathbb{R}^d$. We consider the distributional strategy $\{X_u\}_{u\in\mathcal{V}}$ defined by

$$\Pr[X_u = (r,i)] = \begin{cases} \boldsymbol{u}(r)^2 & \text{if } \operatorname{sign}\boldsymbol{u}(r) = (-1)^i, \\ 0 & \text{otherwise.}\end{cases}$$

Here $\boldsymbol{u}(r)$ denotes the r^{th} coordinate of $\boldsymbol{u}$. In other words, we choose r according to the probability distribution given by $\boldsymbol{u}(1)^2, \ldots, \boldsymbol{u}(d)^2$ and then set i to be 1 or 2 depending on the sign of $\boldsymbol{u}(r)$.

Let us first consider the case that π is the identity permutation. The square of the Hellinger distance of X_u and X_v can be upper bounded by

$$\begin{aligned}\mathrm{H}^2(X_u, X_v) &= \tfrac{1}{2}\sum_{\substack{r\in[n],\\ \operatorname{sign}\boldsymbol{u}(r)=\operatorname{sign}\boldsymbol{v}(r)}} \Big(\boldsymbol{u}(r)-\boldsymbol{v}(r)\Big)^2 + \tfrac{1}{2}\sum_{\substack{r\in[n],\\ \operatorname{sign}\boldsymbol{u}(r)\neq\operatorname{sign}\boldsymbol{v}(r)}} \boldsymbol{u}(r)^2 + \boldsymbol{v}(r)^2 \\ &\leqslant \tfrac{1}{2}\|\boldsymbol{u}-\boldsymbol{v}\|^2 = \tfrac{1}{2}\|(\boldsymbol{u}_1-\boldsymbol{v}_1)-(\boldsymbol{u}_2-\boldsymbol{v}_2)\|^2\,.\end{aligned}$$

The triangle inequality implies $\mathrm{H}^2(X_u, X_v) \leqslant \frac{1}{2}(\|\boldsymbol{u}_1 - \boldsymbol{v}_1\| + \|\boldsymbol{u}_2 - \boldsymbol{v}_2\|)^2 \leqslant \|\boldsymbol{u}_1 - \boldsymbol{v}_1\|^2 + \|\boldsymbol{u}_2 - \boldsymbol{v}_2\|^2$, as desired. If π is the permutation $\pi(i) = 3 - i$, then the same calculation as before shows $\mathrm{H}^2(\pi.X_u, X_v) \leqslant \frac{1}{2}\|\boldsymbol{u}+\boldsymbol{v}\|^2$. Again, the triangle inequality implies $\mathrm{H}^2(\pi.X_u, X_v) \leqslant \|\boldsymbol{u}_1 - \boldsymbol{v}_2\|^2 + \|\boldsymbol{u}_2 - \boldsymbol{v}_1\|^2$. Since there are no other permutations for $k = 2$, the proof is complete. □

5.2. Proof of Lemma 4.6

Lemma 4.6 (Restated)**.** *Let $t \geqslant 1$ and let $\{\boldsymbol{u}_i\}_{u\in\mathcal{V},i\in[k]}$ be a feasible solution of the SDP in Figure 1. Then, there exists a distributional strategy $\{X_u\}_{u\in\mathcal{V}}$ such that for every (u, v, π),*

$$\mathrm{H}^2(\pi.X_u, X_v) \leqslant O(t)\cdot \textstyle\sum_{i\in[k]} \tfrac{1}{2}\|\boldsymbol{u}_i - \boldsymbol{v}_{\pi(i)}\|^2 + k\cdot 2^{-t}\,. \qquad (15)$$

Let $\{\boldsymbol{u}_i\}_{u\in\mathcal{V},i\in[k]}$ be a feasible SDP solution with $\boldsymbol{u}_i \in \mathbb{R}^d$. The distributional strategy $\{X_u\}_{u\in\mathcal{V}}$ we construct will consist of distributions $X_u = (R_u, L_u)$ over $\Omega\times[k]$ with $\Omega = \mathbb{R}^d$.

The basic building blocks of our constructions are distributions of the following kind: For $\boldsymbol{w} \in \mathbb{R}^d$, let $D_{\boldsymbol{w}}$ denote the distribution over $\mathbb{R}^d$ whose density at $\boldsymbol{x} \in \mathbb{R}^d$ is equal to

$$\gamma_{\sigma,\boldsymbol{w}}(\boldsymbol{x}) \overset{\text{def}}{=} \frac{1}{(2\pi)^{n/2}\sigma^n}\cdot\exp\left(-\tfrac{1}{2}\left\|\tfrac{1}{\sigma}(\boldsymbol{x}-\boldsymbol{w})\right\|_2^2\right).$$

Here σ is a parameter that we choose as $\sigma = 1/C\sqrt{t}$ for some large enough constant $C > 0$. The distribution $D_{\boldsymbol{w}}$ is the standard d-dimensional Gaussian distribution translated by the vector $\boldsymbol{w}$ and scaled by the factor σ.

For $u \in \mathcal{V}$ and $i \in [k]$, we define a distribution $X_u^{(i)} = (R_u^{(i)}, L_u^{(i)})$ over $\mathbb{R}^d\times[k]$ as follows. The first component $R_u^{(i)}$ is distributed according to $D_{\tilde{\boldsymbol{u}}_i}$, where we denote by $\tilde{\boldsymbol{w}}$ the unit vector $\frac{1}{\|\boldsymbol{w}\|}\boldsymbol{w}$ in direction $\boldsymbol{w}$. The second component $L_u^{(i)}$ is equal to the index $j \in [k]$ for which the projection of $R_u^{(i)}$ on $\tilde{\boldsymbol{u}}_j$ is largest. Formally, the density function f of $X_u^{(i)}$ is

$$f(\boldsymbol{x}, j) = \begin{cases}\gamma_{\sigma,\tilde{\boldsymbol{u}}_i}(\boldsymbol{x}) & \text{if } \langle \boldsymbol{x}, \tilde{\boldsymbol{u}}_j\rangle > \max_{h\in[k]\setminus\{j\}}\langle\boldsymbol{x},\tilde{\boldsymbol{u}}_h\rangle, \\ 0 & \text{otherwise.}\end{cases}$$

Finally, we define the distribution X_u as the convex combination of the distributions $X_u^{(1)}, \ldots, X_u^{(k)}$ with coefficients $\|\boldsymbol{u}_1\|^2, \ldots, \|\boldsymbol{u}_k\|^2$.

The following claim shows that we can upper bound the Hellinger distance of $D_{\boldsymbol{u}}$ and $D_{\boldsymbol{v}}$ in terms of the Euclidean distance of $\boldsymbol{u}$ and $\boldsymbol{v}$.

Claim 5.1. *For any two vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^d$,*

$$\mathrm{H}^2(D_{\boldsymbol{u}}, D_{\boldsymbol{v}}) \leqslant \tfrac{1}{\sigma^2}\|\boldsymbol{u}-\boldsymbol{v}\|^2\,. \qquad (16)$$

Proof. For any $\boldsymbol{x} \in \mathbb{R}^d$, one gets

$$\begin{aligned}\sqrt{\gamma_{\sigma,\boldsymbol{u}}(\boldsymbol{x})\cdot\gamma_{\sigma,\boldsymbol{v}}(\boldsymbol{x})} &= \frac{1}{(2\pi)^{n/2}\sigma^n}\cdot e^{-\frac{1}{4}\|\frac{1}{\sigma}(\boldsymbol{x}-\boldsymbol{u})\|^2-\frac{1}{4}\|\frac{1}{\sigma}(\boldsymbol{x}-\boldsymbol{v})\|^2} \\ &= \frac{1}{(2\pi)^{n/2}\sigma^n}\cdot e^{-\frac{1}{8}\|\frac{1}{\sigma}(2\boldsymbol{x}-(\boldsymbol{u}+\boldsymbol{v}))\|^2-\frac{1}{8}\|\frac{1}{\sigma}(\boldsymbol{u}-\boldsymbol{v})\|^2} \\ &= \exp\left(-\tfrac{1}{8}\|\tfrac{1}{\sigma}(\boldsymbol{u}-\boldsymbol{v})\|^2\right)\cdot\gamma_{\sigma,\frac{1}{2}(\boldsymbol{u}+\boldsymbol{v})}(\boldsymbol{x}) \\ &\geqslant (1-\tfrac{1}{\sigma^2}\|\boldsymbol{u}-\boldsymbol{v}\|^2)\cdot\gamma_{\sigma,\frac{1}{2}(\boldsymbol{u}+\boldsymbol{v})}(\boldsymbol{x})\,, \qquad (17)\end{aligned}$$

where the second equality follows by the parallelogram law, $\|\boldsymbol{a}\|^2 + \|\boldsymbol{b}\|^2 = \frac{1}{2}\|\boldsymbol{a}+\boldsymbol{b}\|^2 + \frac{1}{2}\|\boldsymbol{a}-\boldsymbol{b}\|^2$, and the last step follows from the fact that $1 + x \leqslant e^x$ for all $x \in \mathbb{R}$.

Thus, the Hellinger distance of $D_{\boldsymbol{u}}$ and $D_{\boldsymbol{v}}$ satisfies

$$\mathrm{H}^2(D_{\boldsymbol{u}}, D_{\boldsymbol{v}}) = 1 - \int \sqrt{\gamma_{\sigma,\boldsymbol{u}}\gamma_{\sigma,\boldsymbol{v}}}$$
$$\overset{(17)}{\leqslant} 1 - \left(1 - \tfrac{1}{\sigma^2}\|\boldsymbol{u}-\boldsymbol{v}\|^2\right) \int \gamma_{\sigma,\frac{1}{2}(\boldsymbol{u}+\boldsymbol{v})} = \tfrac{1}{\sigma^2}\|\boldsymbol{u}-\boldsymbol{v}\|^2 .$$

□

Using standard tail bounds for the Gaussian distribution, we can derive a bound similar to (16) for the Hellinger distance of $\pi.X_u^{(i)}$ and $X_v^{(\pi(i))}$.

Claim 5.2. *For every $i \in [k]$ and every permutation π of $[k]$,*

$$\mathrm{H}^2(\pi.X_u^{(i)}, X_v^{(\pi(i))}) \leqslant O(t) \cdot \|\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{v}}_{\pi(i)}\|^2 + k \cdot 2^{-t} .$$

Proof. Let $j = \pi(i)$. By the triangle inequality for the Hellinger distance (Lemma 3.3) and the inequality $(a+b+c)^2 \leqslant 3(a^2+b^2+c^2)$, we have

$$\mathrm{H}^2(\pi.X_u^{(i)}, X_v^{(j)}) \leqslant 3\Big(\mathrm{H}^2(\pi.X_u^{(i)}, (R_u^{(i)}, j)) + \mathrm{H}^2((R_u^{(i)}, j), (R_v^{(j)}, j)) + \mathrm{H}^2((R_v^{(j)}, j), X_v^{(j)})\Big) . \qquad (18)$$

Claim 5.1 implies for the second term on the rhs of (18),

$$\mathrm{H}^2((R_u^{(i)}, j), (R_v^{(j)}, j)) = \mathrm{H}^2(D_{\tilde{\boldsymbol{u}}_i}, D_{\tilde{\boldsymbol{v}}_j}) \leqslant O(t) \cdot \|\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{v}}_j\|^2 .$$

We bound the first term on the rhs of (18) by the corresponding statistical distance,

$$\mathrm{H}^2(\pi.X_u^{(i)}, (R_u^{(i)}, j)) \leqslant \Delta(\pi.X_u^{(i)}, (R_u^{(i)}, j)) = \Pr[\pi(L_u^{(i)}) \neq j]$$
$$= \sum_{h\in[k]\setminus\{i\}} \Pr[L_u^{(i)} = h] \leqslant \sum_{h\in[k]\setminus\{i\}} \Pr_{\boldsymbol{x}\sim D_{\tilde{\boldsymbol{u}}_i}} [\langle \boldsymbol{x}, \tilde{\boldsymbol{u}}_h\rangle > \langle \boldsymbol{x}, \tilde{\boldsymbol{u}}_i\rangle] .$$

We can write $\boldsymbol{x}$ as $\tilde{\boldsymbol{u}}_i + \sigma\boldsymbol{g}$, where $\boldsymbol{g}$ is a standard Gaussian vector. Hence, the event $[\langle \boldsymbol{x}, \tilde{\boldsymbol{u}}_h\rangle > \langle \boldsymbol{x}, \tilde{\boldsymbol{u}}_i\rangle] = [\langle \boldsymbol{g}, \tilde{\boldsymbol{u}}_h - \tilde{\boldsymbol{u}}_i\rangle > 1/\sigma]$. The inner product $\langle \boldsymbol{g}, \tilde{\boldsymbol{u}}_h - \tilde{\boldsymbol{u}}_i\rangle$ has a Gaussian distribution with mean 0 and standard deviation $\sqrt{2}$. Therefore, by standard estimates of the tail of the Gaussian distribution, the probability of this event is at most $e^{-\frac{1}{16\sigma^2}}$. Thus, the first term on the right-hand side of (18) contributes at most $3k \cdot e^{-\frac{1}{16\sigma^2}} \leqslant 1/2 \cdot k \cdot 2^{-t}$. The same is true for the third term in (18). The claim follows. □

Using the previous two claims, we can now show the bound (15) on the squared Hellinger distance $\mathrm{H}^2(\pi.X_u, X_v)$. Since $\pi.X_u$ and X_v are convex combinations of the distributions $\pi.X_u^{(1)}, \ldots, \pi.X_u^{(k)}$ and $X_v^{(1)}, \ldots, X_v^{(k)}$, respectively, Lemma 3.4 implies that $\mathrm{H}^2(\pi.X_u, X_v)$ is at most

$$\sum_{i\in[k]} \|\boldsymbol{u}_i\| \cdot \|\boldsymbol{v}_{\pi(i)}\| \cdot \mathrm{H}^2(\pi.X_u^{(i)}, X_v^{(\pi(i))}) + \tfrac{1}{2} \sum_{i\in[k]} (\|\boldsymbol{u}_i\| - \|\boldsymbol{v}_{\pi(i)}\|)^2 . \quad (19)$$

The second sum in (19) contributes at most $\sum_{i\in[k]} \|\boldsymbol{u}_i - \boldsymbol{v}_{\pi(i)}\|^2$, because for any two vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^d$, $(\|\boldsymbol{u}\| - \|\boldsymbol{v}\|)^2 \leqslant \|\boldsymbol{u}-\boldsymbol{v}\|^2$ (triangle inequality). On the other hand, Claim 5.2 allows us to bound the first sum in (19) by

$$\sum_{i\in[k]} \|\boldsymbol{u}_i\| \cdot \|\boldsymbol{v}_{\pi(i)}\| \cdot \Big(O(t) \cdot \|\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{v}}_{\pi(i)}\|^2 + k \cdot 2^{-t}\Big)$$
$$\leqslant O(t) \sum_{i\in[k]} \|\boldsymbol{u}_i\| \cdot \|\boldsymbol{v}_{\pi(i)}\| \cdot \|\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{v}}_{\pi(i)}\|^2 \; + \; k \cdot 2^{-t}$$
$$\leqslant O(t) \sum_{i\in[k]} \|\boldsymbol{u}_i - \boldsymbol{v}_{\pi(i)}\|^2 \; + \; k \cdot 2^{-t} ,$$

where we used in the first step that $(\sum_{i\in[k]} \|\boldsymbol{u}_i\| \cdot \|\boldsymbol{v}_{\pi(i)}\|)^2 \leqslant \sum_{i\in[k]} \|\boldsymbol{u}_i\|^2 \cdot \sum_{i\in[k]} \|\boldsymbol{v}_{\pi(i)}\|^2 = 1$ (Cauchy–Schwarz) and in the second step $\|\boldsymbol{u}_i\| \cdot \|\boldsymbol{v}_{\pi(i)}\| \cdot \|\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{v}}_{\pi(i)}\|^2 = 2\|\boldsymbol{u}_i\| \cdot \|\boldsymbol{v}_{\pi(i)}\| - 2\langle \boldsymbol{u}_i, \boldsymbol{v}_{\pi(i)}\rangle \leqslant \|\boldsymbol{u}_i\|^2 + \|\boldsymbol{v}_{\pi(i)}\|^2 - 2\langle \boldsymbol{u}_i, \boldsymbol{v}_{\pi(i)}\rangle = \|\boldsymbol{u}_i - \boldsymbol{v}_{\pi(i)}\|^2$ (AM–GM inequality).The proof of Lemma 4.6 is complete. □

6. Conclusions and Open Problems

Our results show that for unique games, the value that the semidefinite program really captures is not the integral value of the game but rather the amortized value under many parallel repetitions, i.e., the value $\overline{\mathsf{val}}(G) = \lim_{\ell\to\infty} \mathsf{val}(G^\ell)^{1/\ell}$. If Khot's unique game conjecture [15] is true then this means that the amortized value can be much easier to approximate than the original value of the game. We find this quite surprising, as in general computing the amortized value of even very simple finite games is considered a very hard problem.

Can one get rid of the $\log(1/\delta)$ term in Theorem 1.1? We conjecture that this should be possible by using a more careful construction of the distributions R_u, although some subtleties seem to arise. Another interesting open question is whether strong parallel repetition holds for unique games with entanglement. Although the authors disagree on the answer to this question, it seems that some important insight on it can be obtained by combining our techniques with those of [14]. A more general question is to find more applications of the interplay between the Hellinger and statistical distance. One such application was recently found by Kindler et al. [18] who used Raz's ideas to construct more efficient foams in $\mathbb{R}^d$.

A further consequence of our work is that Khot's unique games conjecture [15] is equivalent to the following, a priori stronger hypothesis: for every $\varepsilon > 0$, there exists an alphabet size k such that given a unique game G, it is **NP**-hard to distinguish between the case that (1) $\mathsf{val}(G) \geqslant 1 - \varepsilon$ and (2) for every distributional strategy $\{X_u\}_{u\in\mathcal{V}}$, the expected squared Hellinger distance $\mathsf{E}_{(u,v,\pi)\sim G}\, \mathrm{H}^2(\pi.X_u, X_v) \geqslant 1 - \varepsilon$. We refer to the full version of the paper for further details on this connection to the unique games conjecture.

Acknowledgments: We thank Guy Kindler for useful discussions, and Thomas Vidick and Nisheeth Vishnoi for helpful comments. We also thank Prasad Raghavendra and Nisheeth Vishnoi for letting us know about the integrality gaps for unique games implied by [17] and [16].

References

[1] S. Aaronson. The ten most annoying questions in quantum computing, 2006. Available at `http://scottaaronson.com/blog/?p=112`.

[2] K. Azimian and M. Szegedy. Parallel repetition of the odd cycle game. In *LATIN*, pages 676–686, 2008.

[3] J. S. Bell. On the Einstein-Podolsky-Rosen paradox. *Physics*, 1(3):195–290, 1964.

[4] A. Broder. On the resemblance and containment of documents. In *Proc. Compression and Complexity of Sequences*, pages 21–29. IEEE, 1997.

[5] M. Charikar, K. Makarychev, and Y. Makarychev. Near-optimal algorithms for unique games. In *Proc.* 38*th STOC*, pages 205–214. ACM, 2006.

[6] M. S. Charikar. Similarity estimation techniques from rounding algorithms. In *Proc.* 34*th STOC*, pages 380–388. ACM, 2002.

[7] E. Chlamtac, K. Makarychev, and Y. Makarychev. How to play unique games using embeddings. In *Proc.* 47*th FOCS*, pages 687–696. IEEE, 2006.

[8] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.*, 23(15):880–884, Oct 1969.

[9] U. Feige, G. Kindler, and R. O'Donnell. Understanding parallel repetition requires understanding foams. In *IEEE Conference on Computational Complexity*, pages 179–192, 2007.

[10] U. Feige and L. Lovász. Two-prover one-round proof systems: Their power and their problems. In *Proc.* 24*th STOC*, pages 733–741. ACM, 1992.

[11] A. Gibbs and F. Su. On choosing and bounding probability metrics. *International Statistical Review*, 70(3):419–435, 2002. Available as eprint arXiv:math/0209021v1.

[12] J. Håstad. Some optimal inapproximability results. *Journal of the ACM*, 48(4):798–859, 2001. Preliminary version in STOC'97.

[13] T. Holenstein. Parallel repetition: simplifications and the no-signaling case. In *Proc.* 39*th STOC*, pages 411–419. ACM, 2007.

[14] J. Kempe, O. Regev, and B. Toner. Unique games with entangled provers are easy. In *Proc.* 49*th FOCS*. IEEE, 2008. Available as eprint arXiv:0710.0655.

[15] S. Khot. On the power of unique 2-prover 1-round games. In *Proc.* 34*th STOC*, pages 767–775. ACM, 2002.

[16] S. Khot, G. Kindler, E. Mossel, and R. O'Donnell. Optimal inapproximability results for MAX-CUT and other 2-variable CSPs? *SIAM J. Comput*, 37(1):319–357, 2007. Preliminary version in FOCS' 04.

[17] S. Khot and N. K. Vishnoi. The unique games conjecture, integrality gap for cut problems and embeddability of negative type metrics into ℓ_1. In *Proc.* 46*th FOCS*, pages 53–62. IEEE, 2005.

[18] G. Kindler, R. O'Donnell, A. Rao, and A. Wigderson. Rounding schemes and cubical tilings with sphere-like surface area. In *Proc.* 49*th FOCS*. IEEE, 2008.

[19] J. M. Kleinberg and É. Tardos. Approximation algorithms for classification problems with pairwise relationships: metric labeling and markov random fields. *J. ACM*, 49(5):616–639, 2002. Preliminary version in STOC'99.

[20] C. Kraft. *Some Conditions for Consistency and Uniform Consistency of Statistical Procedures*. University of California Press, 1955.

[21] R. Mittal and M. Szegedy. Product rules in semidefinite programming. In *Proc.* 16*th FCT*, pages 435–445. Springer, 2007.

[22] D. Pollard. *A user's guide to measure theoretic probability*. Cambridge University Press, 2002.

[23] A. Rao. Parallel repetition in projection games and a concentration bound. In *Proc.* 40*th STOC*, pages 1–10. ACM, 2008. Technical report available on ECCC as report TR08-013.

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

[25] R. Raz. A counterexample to strong parallel repetition. In *Proc.* 49*th FOCS*. IEEE, 2008. Technical report available on ECCC as report TR08-018.

A Further Proofs

A.1 Proof of Lemma 4.3

Let $\{X_u\}_{u\in\mathcal{V}}$ be a distributional strategy with $X_u = (R_u, L_u)$ distributed over $\Omega \times [k]$ such that $\mathsf{E}_{(u,v,\pi)\sim G}\,\Delta(\pi.X_u, X_v) = \eta$. Our goal is to show $\mathsf{val}(G) \geqslant 1 - 2\eta$. Let $h_u \colon \Omega \to [k]$ be the function such that $L_u = h_u(R_u)$. An easy calculation shows that for every triple (u, v, π) there are two disjoint sets $B_1, B_2 \subseteq \Omega$ such that $h_v(r) \neq \pi(h_u(r))$ if and only if $r \in B_1 \cup B_2$, and

$$\Delta(\pi.X_u, X_v) = \Delta(R_u, R_v) + \Pr[R_u \in B_1] + \Pr[R_v \in B_2]\,. \quad (20)$$

Let Z be a random variable uniformly distributed in $[0, 1]$. The correlated sampling lemma yields a collection of functions $\{r_u \colon [0,1] \to \Omega\}_{u\in\mathcal{V}}$ such that $r_u(Z)$ is distributed according to R_u and $\Pr[r_u(Z) \neq r_v(Z)] = 2\Delta(R_u, R_v)/(1 + \Delta(R_u, R_v))$. Define $l_u \colon [0, 1] \to [k]$ as $l_u(Z) = h_u(r_u(Z))$. To derive a lower bound on the value of G, we estimate the probability

$$\begin{aligned}
&\Pr[l_v(Z) \neq \pi(l_u(Z))] \\
&\leqslant \Pr[r_u(Z) \neq r_v(Z)] + \Pr[r_u(Z) \in B_1] + \Pr[r_v(Z) \in B_2] \\
&\overset{(20)}{=} \frac{2\Delta(R_u, R_v)}{1 + \Delta(R_u, R_v)} + \Delta(\pi.X_u, X_v) - \Delta(R_u, R_v) \\
&\leqslant 2\Delta(\pi.X_u, X_v)\,, \qquad (21)
\end{aligned}$$

where we use in the last step $0 \leqslant \Delta(R_u, R_v) \leqslant \Delta(\pi.X_u, X_v)$. Hence, the value of G is at least

$$\begin{aligned} \mathsf{val}(G) &\geqslant \mathop{\mathsf{E}}_{(u,v,\pi)\sim G} \Pr[l_v(Z) = \pi(l_u(Z))] \\ &\overset{(21)}{\geqslant} 1 - \mathop{\mathsf{E}}_{(u,v,\pi)\sim G} 2\Delta(\pi.X_u, X_v) = 1 - 2\eta\,. \quad \square \end{aligned}$$

Remark A.1. The above lower bound $\mathsf{val}(G) \geqslant 1 - 2\eta$ is non-trivial only for $\eta < 1/2$. Using a more precise version of the correlated sampling lemma, one can improve the lower bound to $\mathsf{val}(G) \geqslant 1-\eta/1+\eta$. This bound gives a non-trivial guarantee whenever $\eta < 1$. Details will be presented in the full version of the paper.