A Parallel Repetition Theorem for the GHZ Game

Justin Holmgren[*]   Ran Raz[†]

August 11, 2020

Abstract

We prove that parallel repetition of the (3-player) GHZ game reduces the value of the game polynomially fast to 0. That is, the value of the GHZ game repeated in parallel t times is at most $t^{-\Omega(1)}$. Previously, only a bound of $\approx \frac{1}{\alpha(t)}$, where α is the inverse Ackermann function, was known [Ver96].

The GHZ game was recently identified by Dinur, Harsha, Venkat and Yuen as a multi-player game where all existing techniques for proving strong bounds on the value of the parallel repetition of the game fail. Indeed, to prove our result we use a completely new proof technique. Dinur, Harsha, Venkat and Yuen speculated that progress on bounding the value of the parallel repetition of the GHZ game may lead to further progress on the general question of parallel repetition of multi-player games. They suggested that the strong correlations present in the GHZ question distribution represent the "hardest instance" of the multi-player parallel repetition problem [DHVY17].

Another motivation for studying the parallel repetition of the GHZ game comes from the field of quantum information. The GHZ game, first introduced by Greenberger, Horne and Zeilinger [GHZ89], is a central game in the study of quantum entanglement and has been studied in numerous works. For example, it is used for testing quantum entanglement and for device-independent quantum cryptography. In such applications a game is typically repeated to reduce the probability of error, and hence bounds on the value of the parallel repetition of the game may be useful.

Contents

1   Introduction

In a k-player game, players are given correlated "questions" $(q_1, \ldots, q_k)$ sampled from a distribution Q and must produce corresponding "answers" $(a_1, \ldots, a_k)$ such that $(q_1, \ldots, q_k, a_1, \ldots, a_k)$ satisfy a fixed predicate π. Crucially, the players are not allowed to communicate amongst themselves after receiving their questions (but they may agree upon a strategy beforehand). The value of the game is the probability with which the players can win with an optimal strategy. Multi-player games play a central role in theoretical computer science due to their intimate connection with multi-prover interactive proofs (MIPs) [BGKW88], hardness of approximation [FGL$^+$91], communication complexity [PRW97, BJKS04], and the EPR paradox and non-local games [EPR35, CHTW04]

One basic operation on multi-player games is parallel repetition. In the t-wise parallel repetition of a game, question tuples $(q_1^{(i)}, \ldots, q_k^{(i)})$ are sampled independently for $i \in [t]$. The jth player is given $(q_j^{(1)}, \ldots, q_j^{(t)})$, and is required to produce $(a_j^{(1)}, \ldots, a_j^{(t)})$. The players win if for every $i \in [t]$, $(a_1^{(i)}, \ldots, a_k^{(i)})$ is a winning answer for questions $(q_1^{(i)}, \ldots, q_k^{(i)})$. Parallel repetition was first proposed in [FRS94] as an intuitive attempt to reduce the value of a game from ϵ to ϵ^t, but in general this is not what happens [For89, Fei91, FV96, Raz11]. The actual effect is far more subtle and a summary of some of the known results is given in Table 1.

	Two-player games	$\geq$ 3-player games
Classical	$\exp(-\Omega(t))$ [Raz98]	$O\left(\frac{1}{\alpha(t)}\right)$ [Ver96]
Entangled	$t^{-\Omega(1)}$ [Yue16]	$O(1)$ (trivial)
Non-Signaling	$\exp\left(-\Omega\left(t\right)\right)$ [Hol09]	$\Omega(1)$ [HY19]

Table 1: Known bounds on the worst-case (slowest) decay for various values of the t-wise parallel repetition of a non-trivial game. α denotes the inverse Ackermann function.

Much less is known about games with three or more players than about two-player games. Only very weak bounds are known on how t-wise parallel repetition decreases the value of a three-player game (as a function of t). There is a similar gap in our understanding when players are allowed to share entangled state; in fact, no bounds here are known whatsoever in the three-player case. If players are more generally allowed to use any no-signaling strategy, then there are in fact counterexamples (lower bounds) showing that parallel repetition may utterly fail to reduce the (no-signaling) value of a three-player game.

1.1   The GHZ Game

The GHZ game, which we will denote by $\mathcal{G}_{\mathsf{GHZ}}$, is a three-player game with query distribution Q_{GHZ} that is uniform on $\{x \in \mathbb{F}_2^3 : x_1 + x_2 + x_3 = 0\}$. To win, players are required on input (x_1, x_2, x_3) to produce (y_1, y_2, y_3) such that $y_1 \oplus y_2 \oplus y_3 = x_1 \vee x_2 \vee x_3$. It is easily verified that the value of $\mathcal{G}_{\mathsf{GHZ}}$ is $3/4$.

Dinur et al. [DHVY17] identified the GHZ game as a simple example of a game for which we do not know exponential decay bounds, writing

> *"We suspect that progress on bounding the value of the parallel repetition of the GHZ game will lead to further progress on the general question."*

and

> *"We believe that the strong correlations present in the GHZ question distribution represent the "hardest instance" of the multiplayer parallel repetition problem. Existing techniques from the two-player case (which we leverage in this paper) appear to be incapable of analyzing games with question distributions with such strong correlations."*

The GHZ game also plays an important role in quantum information theory and in particular in entanglement testing and device-independent quantum cryptography. Its salient properties are that it is an

XOR game for which quantum (entangled) players can play perfectly, but classical players can win only with probability strictly less than 1 [MS13]. No such *two*-player game is known. Moreover, the GHZ game has the so called, self testing property, that all quantum strategies that achieve value 1 are essentially equivalent. This property is important for entanglement testing and device-independent quantum cryptography.

Prior to our work, the best known parallel repetition bound for the GHZ game was due to Verbitsky [Ver96], who observed a connection between parallel repetition and the density Hales-Jewett theorem from Ramsey theory [FK91]. Using modern quantitative versions of this theorem [Pol12], Verbitsky's result implies a bound of approximately $\frac{1}{\alpha(t)}$, where α is the inverse Ackermann function.

We prove a bound of $t^{-\Omega(1)}$.

2 Technical Overview

To prove our parallel repetition theorem for the GHZ game we show that for an arbitrary strategy, even if we condition on that strategy winning in several coordinates $i_1, \ldots, i_m$, there still exists some coordinate in which that strategy loses with significant probability. We consider the finer-grained event that also specifies specific queries and answers in coordinates $i_1, \ldots, i_m$, and abstract it out as a sufficiently dense product event E over the three players' inputs.

Given an arbitrary product event E that occurs with sufficiently high probability, we show that some coordinate of $\tilde{P} \stackrel{\mathsf{def}}{=} P|E$ is hard. We do this in three high-level steps:

1. We first prove this for the simpler case in which E is an affine subspace of $\mathbb{F}_2^{3 \times n}$. In fact, we show in this case that *many* coordinates of $\tilde{P}$ are hard.

2. We then prove that when E is arbitrary, $\tilde{P}$ can be written as a convex combination of components $\tilde{P}|\mathcal{W}$, where $\mathcal{W}$ is a large affine subspace, with most such components "indistinguishable" from $P|\mathcal{W}$. Specifically, our main requirement is that for all sufficiently compressing linear functions ϕ on $\mathcal{W}$, the KL divergence of $\phi(\tilde{X})$ from $\phi(X)$ is small, where we sample $\tilde{X} \leftarrow \tilde{P}|\mathcal{W}$ and $X \leftarrow P|\mathcal{W}$.

3. With this notion of indistinguishability, we prove that if $\tilde{P}|\mathcal{W}$ is indistinguishable from $P|\mathcal{W}$, then the GHZ game (or any game with a constant-sized answer alphabet) is roughly as hard in every coordinate with query distribution $\tilde{P}|\mathcal{W}$ as with $P|\mathcal{W}$.

We conclude that for many coordinates i, there is a significant portion of $\tilde{P}$ for which the ith coordinate is hard. We emphasize that unlike all previous parallel repetition bounds, our proof does *not* construct a local embedding of Q_{GHZ} into $\tilde{P}$ for general E.

Local Embeddability in Affine Subspaces We first show that if E is any affine subspace of sufficiently low codimension m in $\mathbb{F}_2^{3 \times n}$, then there exist many coordinates $i \in [n]$ for which Q_{GHZ} is locally embeddable in the ith coordinate of the conditional distribution $\tilde{P}$. In fact, it will suffice for us to consider only affine "power" subspaces, i.e. of the form $w + \mathcal{V}^3$ for some linear subspace $\mathcal{V}$ in $\mathbb{F}_2^n$ and vector $w \in \mathbb{F}_2^{3 \times n}$. Let $X^1, \ldots, X^n \in \mathbb{F}_2^3$ denote the queries in each of the n repetitions.

Our observation is that when E is affine there exists a subset of coordinates $S \subseteq [n]$ with $|S| \geq 2$ such that for any $i' \in S$, E depends on $(X^i)_{i \in S}$ only via the differences $(X^{i'} - X^i)_{i' \in S \setminus \{i\}}$. Indeed, if $E = E_1 \times E_2 \times E_3$ and if each E_j is given by an affine equation $(X_j^1, \ldots, X_j^n) \cdot A = b_j$ for a sufficiently "skinny" matrix A, then by the pigeonhole principle there must exist two distinct subset row-sums of A with equal values. By considering the symmetric difference of these subsets, and using the fact that we are working over $\mathbb{F}_2$, there is a set $S \subseteq [n]$ such that the S-subset row-sum of A is 0. Thus the value of $(X_j^1, \ldots, X_j^n) \cdot A$ is unchanged if X_j^i is subtracted from $X_j^{i'}$ for every $i \in S$.

As a result, the players can all sample $(X^{i'} - X^i)_{i' \in S \setminus \{i\}}$ and $(X^{i'})_{i' \notin S}$, which are independent of X^i, using shared randomness. On input X_j^i, the jth player can locally compute $(X_j^{i'})_{i' \in S}$ from X_j^i and $(X^{i'} - X^i)$.

Pseudo-Affine Decompositions   At a high level, we next show that if E is an arbitrary product event (with sufficient probability mass) then $\tilde{P}$ has a "pseudo-affine decomposition". That is, there is a partition Π of $(\mathbb{F}_2^n)^3$ into affine subspaces such that if $\mathcal{W}$ is a random part of Π (as weighted by $\tilde{P}$), then any strategy for $\tilde{P}|\mathcal{W}$ can be extended to a strategy for $P|\mathcal{W}$ that is similarly successful in expectation.

To construct Π, we prove the following sufficient conditions for Π to be a pseudo-affine decomposition:

- When $\mathcal{W}$ is a random part of Π (as weighted by $\tilde{P}$), the distributions $\tilde{P}|\mathcal{W}$ and $P|\mathcal{W}$ are indistinguishable to all sufficiently compressing linear distinguishers. That is, if $\mathcal{W}$ is an affine shift of $\mathcal{V}^3$, then for all subspaces $\mathcal{U} \leq \mathcal{V}$ of sufficiently small co-dimension, the distributions $\tilde{P}|\mathcal{W}$ and $P|\mathcal{W}$ are statistically close modulo $\mathcal{U}^3$.

- Each part $\mathcal{W}$ of Π is in fact an affine shift of a *product* space $\mathcal{V}^3$ for some linear space $\mathcal{V}$.

We construct Π satisfying these conditions iteratively. Starting with the singleton partition, as long as a random part $\mathcal{W}$ of Π has some subspace $\mathcal{U}$ for which $\tilde{P}|\mathcal{W}$ and $P|\mathcal{W}$ are distributed differently mod $\mathcal{U}^3$, we replace each part $\mathcal{W}$ of Π by all the affine shifts of $\mathcal{U}^3$ in $\mathcal{W}$. We show that this process cannot be repeated too many times when E has sufficient density.

Pseudorandomness Preserves Hardness   The high-level reason these conditions suffice is because for any strategy $f = f_1 \times f_2 \times f_3$, they enable us to refine Π to a partition Π'_f such that when X is sampled from $\tilde{P}|\mathcal{W}'$ for a random part $\mathcal{W}'$ in Π'_f, the distribution of $f(X)$ is as if X were sampled *uniformly* from $\mathcal{W}' \cap E$ (i.e. with X_1, X_2, and X_3 mutually independent). Moreover, when we construct Π'_f we partition each part $\mathcal{W}$ of Π into all affine shifts of some linear space $\mathcal{U}^3$ where the codimension of $\mathcal{U}^3$ in $\mathcal{W}$ is not too large. Thus the strategy f on $\tilde{P}|\mathcal{W}$ effectively has the players acting as independent (randomized) functions of their inputs modulo $\mathcal{U}$. Such strategies generalize to $P|\mathcal{W}$ by the first property of pseudo-affine decompositions stated above.

To construct Π'_f, we ensure that f_1 is uncorrelated with every affine function on $\tilde{P}|\mathcal{W}'$ when $\mathcal{W}'$ is a random part of Π'_f, and then prove the desired independence by Fourier analysis. We construct Π'_f by iterative refinement of Π. Start by considering a random part $\mathcal{W}$ of Π. Whenever $f(X_1)$ is correlated with an affine $\mathbb{F}_2$-valued function χ, replace $\mathcal{W}$ in Π by $\mathcal{W} \cap \chi^{-1}(0)$ and $\mathcal{W} \cap \chi^{-1}(1)$, and do this in parallel for all parts of Π. We show that this cannot be repeated too many times, and thus we quickly arrive at our desired Π'_F.

3   Preliminaries

In this section we describe some preliminary definitions that are somewhat specific to this work. More standard preliminaries are given in Appendices A and B.

3.1   Set Theory

Definition 3.1. *For any set S, a* partition *of S is a pairwise disjoint set of subsets of S, whose union is all of S.*

If Π is a partition of S and x is an element of S, we write $\Pi(x)$ to denote the (unique) element of Π that contains x.

3.2   Linear Algebra

If U is a linear subspace of V, we write $U \leq V$ rather than $U \subseteq V$ to emphasize that U is a subspace rather than an unstructured subset.

We crucially rely on the Cauchy-Schwarz inequality:

Definition 3.2 (Inner Product Space). *A real inner product space is a vector space V over $\mathbb{R}$ together with an operation $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ satisfying the following axioms for all $x, y, z \in V$:*

- *Symmetry: $\langle x, y \rangle = \langle y, x \rangle$.*

- *Linearity in the first[1] argument: $\langle ax + by, z \rangle = a\langle x, z \rangle + b\langle y, z \rangle$.*

- *Positive Definiteness: $\langle x, x \rangle > 0$ if $x \neq 0$.*

Theorem 3.3 (Cauchy-Schwarz). *In any inner product space, it holds for all vectors u and v that $|\langle u, v \rangle|^2 \leq \langle u, u \rangle \cdot \langle v, v \rangle$.*

3.3 Multi-Player Games

In parallel repetition we often work with Cartesian product sets of the form $\mathcal{X} = (\mathcal{X}_1 \times \cdots \times \mathcal{X}_k)^n$. For these sets, we will use superscripts to index the outer product and subscripts to index the inner product. That is, we view elements x of $\mathcal{X}$ as tuples $(x^1, \ldots, x^n)$, where the ith component of x^j is x_i^j. We will also write x_i to denote the vector $(x_i^1, \ldots, x_i^n)$. If $\{E_i \subseteq \mathcal{X}_i\}_{i \in [k]}$ is a collection of subsets indexed by subscripts, we write $E_1 \times \cdots \times E_k$ or $\prod_{i \in [k]} E_i$ to denote the set $\{x \in \mathcal{X} : \forall i \in [k], x_i \in E_i\}$. Similarly, if $\mathcal{Y}$ is a product set $(\mathcal{Y}_1 \times \cdots \times \mathcal{Y}_k)^m$, we say $f : \mathcal{X} \to \mathcal{Y}$ is a product function $f_1 \times \cdots \times f_k$ if $f(x) = y$ for $y_i = f_i(x_i)$.

Definition 3.4 (Multi-player Games). *A k-player game is a tuple $(\mathcal{X}, \mathcal{Y}, P, W)$, where $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_k$ and $\mathcal{Y} = \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_k$ are finite sets, P is a probability measure on $\mathcal{X}$, and $W : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ is a "winning probability" predicate.*

Definition 3.5 (Parallel Repetition). *Given a k-player game $\mathcal{G} = (\mathcal{X}, \mathcal{Y}, Q, W)$, its n-fold parallel repetition, denoted $\mathcal{G}^n$, is defined as the k-player game $(\mathcal{X}^n, \mathcal{Y}^n, Q^n, W^n)$, where $W^n(x, y) \stackrel{\text{def}}{=} \bigwedge_{j=1}^{n} W(x^j, y^j)$.*

Definition 3.6. *The success probability of a function $f = f_1 \times \cdots f_k : \mathcal{X} \to \mathcal{Y}$ in a k-player game $\mathcal{G} = (\mathcal{X}, \mathcal{Y}, Q, W)$ is*

$$v[f](\mathcal{G}) \stackrel{\text{def}}{=} \Pr_{x \leftarrow Q} \left[W\big((x, f(x)) = 1 \big) \right].$$

Definition 3.7. *The value of a k-player game $\mathcal{G} = (\mathcal{X}, \mathcal{Y}, Q, W)$, denoted $v(\mathcal{G})$, is the maximum, over all functions $f = f_1 \times \cdots \times f_k : \mathcal{X} \to \mathcal{Y}$, of $v[f](\mathcal{G})$.*

Fact 3.8. *Randomized strategies are no better than deterministic strategies.*

Definition 3.9 (Value in j^{th} coordinate). *If $\mathcal{G} = (\mathcal{X}, \mathcal{Y}, Q, W^n)$ is a game (with a product winning predicate), the value of $\mathcal{G}$ in the jth coordinate, denoted $v^j(\mathcal{G})$, is the value of the game $(\mathcal{X}, \mathcal{Y}, Q, W')$, where $W'(x, y) = W(x^i, y^i)$.*

Definition 3.10 (Game with Modified Query Distribution). *If $\mathcal{G} = (\mathcal{X}, \mathcal{Y}, Q, W)$ is a game, and P is a probability measure on $\mathcal{X}$, we write $\mathcal{G}|P$ to denote the game $(\mathcal{X}, \mathcal{Y}, P, W)$.*

4 Key Lemmas

In this section, we give some Fourier-analytic conditions (see Appendix B for the basics of Fourier analysis) that imply independence of random variables under the (parallel repeated) GHZ query distribution.

It will be convenient for us to work with probability distributions in terms of their *densities* (see Appendix A for basic probability definitions and notation).

---

[1]Because of symmetry, this implies also linearity in the second argument, aka bilinearity.

Definition 4.1 (Probability Densities)**.** *If $P : \Omega \to \mathbb{R}$ is a probability distribution with support contained in A, then the* density of P in A *is*

$$\varphi : A \to \mathbb{R}$$
$$x \mapsto |A| \cdot P(x).$$

If A is unspecified, then by default it is taken to be Ω.

Lemma 4.2. *Let $\mathcal{V}$ be a (finite) vector space over $\mathbb{F}_2$, let P be uniform on $\{x \in \mathcal{V}^3 : x_1 + x_2 + x_3 = 0\}$, and let U be uniform on $\mathcal{V}^3$.*

For any subset $E = E_1 \times E_2 \times E_3$ of $\mathcal{V}^3$,

$$P(E) = \sum_{\chi \in \hat{\mathcal{V}}} \prod_{i \in [3]} \hat{1}_{E_i}(\chi) = U(E) \cdot \sum_{\chi \in \hat{\mathcal{V}}} \prod_{i \in [3]} \hat{\varphi}_{E_i}(\chi),$$

where φ_{E_i} denotes the density in $\mathcal{V}$ of the uniform distribution on E_i.

Proof. Let φ_P denote the density in $\mathcal{V}^3$ of P. That is,

$$\varphi_P(x_1, x_2, x_3) = \begin{cases} |\mathcal{V}| & \text{if } x_1 + x_2 + x_3 = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$
\begin{aligned}
P(E) &= \mathop{\mathbb{E}}_{x \leftarrow \mathcal{V}^3} [\varphi_P(x) \cdot 1_E(x)] \\
&= \sum_{\chi \in \widehat{\mathcal{V}^3}} \hat{\varphi}_P(\chi) \cdot \hat{1}_E(\chi). \quad\quad\quad \text{(Plancherel)}
\end{aligned}
\tag{1}
$$

We now compute $\hat{\varphi}_P(\chi)$ and $\hat{1}_E(\chi)$. We start by noting that the dual space $\widehat{\mathcal{V}^3}$ is isomorphic to $\hat{\mathcal{V}}^3$. That is, each character $\chi \in \widehat{\mathcal{V}^3}$ is of the form $\chi(x_1, x_2, x_3) = \chi_1(x_1)\chi_2(x_2)\chi_3(x_3)$ for some (uniquely determined) $\chi_1, \chi_2, \chi_3 \in \hat{\mathcal{V}}$ and conversely, each choice of $\chi_1, \chi_2, \chi_3 \in \hat{\mathcal{V}}$ gives rise to some $\chi \in \widehat{\mathcal{V}^3}$.

The Fourier transform of φ_P is given by

$$\hat{\varphi}_P(\chi_1, \chi_2, \chi_3) = \begin{cases} 1 & \text{if } \chi_1 = \chi_2 = \chi_3 \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Since E is a product event, the Fourier transform of $1_E : \mathcal{V}^3 \to \{0, 1\}$ is given by

$$
\begin{aligned}
\hat{1}_E(\chi_1, \chi_2, \chi_3) &= \prod_{i \in [3]} \hat{1}_{E_i}(\chi_i) \\
&= U(E) \cdot \prod_{i \in [3]} \hat{\varphi}_{E_i}(\chi_i).
\end{aligned}
\tag{3}
$$

Substituting Eqs. (2) and (3) into Eq. (1) concludes the proof of the lemma. $\square$

Corollary 4.3. *With $\mathcal{V}$, P, E, and U as in Lemma 4.2,*

$$|P(E) - U(E)| \leq \sum_{\chi \in \hat{\mathcal{V}} \setminus \{1\}} \prod_{i \in [3]} |\hat{1}_{E_i}(\chi)|,$$

where $1 \in \hat{\mathcal{V}}$ denotes the trivial character.

Proof. For any probability density function φ, we have $\hat{\varphi}(1) = 1$, so

$$|P(E) - U(E)| = U(E) \cdot \left| \frac{P(E)}{U(E)} - 1 \right|$$

$$\leq U(E) \cdot \left| \sum_{\chi \in \hat{\mathcal{V}} \setminus \{1\}} \prod_{i \in [3]} \hat{\varphi}_{E_i}(\chi) \right|$$

$$\leq \sum_{\chi \in \hat{\mathcal{V}} \setminus \{1\}} \prod_{i \in [3]} \left| \hat{1}_{E_i}(\chi) \right|. \qquad \square$$

Lemma 4.4. *Let $\mathcal{V}$ be a (finite) vector space over $\mathbb{F}_2$, let P be uniform on $\{x \in \mathcal{V}^3 : x_1 + x_2 + x_3 = 0\}$, let U be uniform on $\mathcal{V}^3$, and let $X = (X_1, X_2, X_3)$ denote the identity[2] random variable on $\mathcal{V}^3$. Let $Y_i = Y_i(X_i)$ be a $\mathcal{Y}_i$-valued random variable for each $i \in [3]$, let $Y = (Y_1, Y_2, Y_3)$, and let $\mathcal{Y} = \mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{Y}_3$.*

Let $\mathcal{W}$ be a subspace of $\mathcal{V}$. If for all $\chi \in \hat{\mathcal{W}}$,

$$\mathop{\mathbb{E}}_{(x,y) \leftarrow P_{X,Y}} \left[d_{\mathsf{TV}} \left(P_{\chi(X_1)|X \in x + \mathcal{W}^3, Y_1 = y_1}, U_{\chi(X_1)|X \in x + \mathcal{W}^3} \right) \right] \leq \epsilon, \tag{4}$$

then

$$\mathop{\mathbb{E}}_{x \leftarrow P_X} \left[d_{\mathsf{TV}} \left(P_{Y|X \in x + \mathcal{W}^3}, U_{Y|X \in x + \mathcal{W}^3} \right) \right] \leq \epsilon \cdot \sqrt{|\mathcal{Y}_2| \cdot |\mathcal{Y}_3|}.$$

Proof. For $x \in \mathcal{V}^3$, we will write $\bar{x}$ to denote the set $x + \mathcal{W}^3$. Recall that $\mathcal{V}/\mathcal{W}$ denotes the set of all cosets $\{x + \mathcal{W}\}_{x \in \mathcal{V}}$. For every $i \in [3]$, every $\bar{x}_i \in \mathcal{V}/\mathcal{W}$, and every $y_i \in \mathcal{Y}_i$, define $1_{i, \bar{x}_i, y_i} : \bar{x}_i \to \{0, 1\}$ to be the indicator for the set $Y_i^{-1}(y_i) \cap \bar{x}_i$. Define $\varphi_{i, \bar{x}_i, y_i}$ to be the density (in $\bar{x}_i$) of the uniform distribution on $Y_i^{-1}(y_i) \cap \bar{x}_i$. That is,

$$\varphi_{i, \bar{x}_i, y_i} : \bar{x}_i \to \mathbb{R}$$
$$\varphi_{i, \bar{x}_i, y_i}(x_i') = \begin{cases} \frac{|\bar{x}_i|}{|Y_i^{-1}(y_i) \cap \bar{x}_i|} & \text{if } Y_i(x_i') = y_i \\ 0 & \text{otherwise.} \end{cases}$$

$\varphi_{i, \bar{x}_i, y_i}$ is easily seen to be related to $1_{i, \bar{x}_i, y_i}$ as

$$1_{i, \bar{x}_i, y_i} = P_{Y_i | \bar{X}_i = \bar{x}_i}(y_i) \cdot \varphi_{i, \bar{x}_i, y_i}.$$

With this notation, our assumption that Eq. (4) holds (for all $\chi \in \hat{\mathcal{W}}$) is equivalent to assuming that for all $\chi \in \hat{\mathcal{W}} \setminus \{1\}$,

$$\mathop{\mathbb{E}}_{(x,y) \leftarrow P_{X,Y}} \left[\left| \hat{\varphi}_{1, \bar{x}_1, y_1}(\chi) \right| \right] \leq 2\epsilon. \tag{5}$$

This is because for all $\chi \in \hat{\mathcal{W}} \setminus \{1\}$, the distribution $U_{\chi(X_1)|X \in x + \mathcal{W}^3}$ is uniform on $\{\pm 1\}$.

In general for $x \in \text{Supp}(P_X)$, we have (by Corollary 4.3) that for any $y \in \mathcal{Y}$,

$$\left| P_{Y|X \in x + \mathcal{W}^3}(y) - U_{Y|X \in x + \mathcal{W}^3}(y) \right| \leq \sum_{\chi \in \hat{\mathcal{W}} \setminus \{1\}} \prod_{i \in [3]} \left| \hat{1}_{i, \bar{x}_i, y_i}(\chi) \right| \tag{6}$$

because:

- the event $E = \{Y = y\}$ is a product event $E_1 \times E_2 \times E_3$, where each $E_i = \{Y_i = y_i\}$ depends only on X_i or equivalently on $X_i - x_i$,

- the distribution $P_{X-x|\bar{X}=\bar{x}}$ is uniform on $\{(\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3) \in \mathcal{W}^3 : \mathbf{w}_1 + \mathbf{w}_2 + \mathbf{w}_3 = 0\}$, and

- the distribution $U_{X-x|\bar{X}=\bar{x}}$ is uniform on $\{(\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3) \in \mathcal{W}^3\}$.

---

[2]Specifically, with the formalism of random variables as functions on a sample space, we mean that X is the identity function, mapping (x_1, x_2, x_3) to (x_1, x_2, x_3).

Thus we have

$$2 \cdot \mathop{\mathbb{E}}_{x \leftarrow P_X} \left[d_{\mathsf{TV}}\big(P_{Y|X \in x + \mathcal{W}^3}, U_{Y|X \in x + \mathcal{W}^3}\big) \right] = \mathop{\mathbb{E}}_{x \leftarrow P_X} \sum_{y \in \mathcal{Y}} \left| P_{Y|X \in x + \mathcal{W}^3}(y) - U_{Y|X \in x + \mathcal{W}^3}(y) \right|$$

$$\leq \mathop{\mathbb{E}}_{x \leftarrow P_X} \sum_{y \in \mathcal{Y}} \sum_{\chi \neq 1} \prod_{i \in [3]} \left| \hat{1}_{i, \bar{x}_i, y_i}(\chi) \right|$$

$$= \mathop{\mathbb{E}}_{x \leftarrow P_X} \sum_{y \in \mathcal{Y}} \sum_{\chi \neq 1} \prod_{i \in \{2,3\}} \sqrt{\left| \hat{1}_{1, \bar{x}_1, y_1}(\chi) \right| \cdot \hat{1}_{i, \bar{x}_i, y_i}(\chi)^2}.$$

Now, we apply Cauchy-Schwarz on the inner product space whose elements are real-valued functions of (x, y, χ), and where the inner product is defined by $\langle f, g \rangle \stackrel{\mathsf{def}}{=} \mathbb{E}_{x \leftarrow P_X} \sum_{y \in \mathcal{Y}} \sum_{\chi \neq 1} f(x, y, \chi) \cdot g(x, y, \chi)$. This bounds the above by

$$\sqrt{\prod_{i \in \{2,3\}} \left(\mathop{\mathbb{E}}_{x \leftarrow P_X} \sum_{y \in \mathcal{Y}} \sum_{\chi \neq 1} \left| \hat{1}_{1, \bar{x}_1, y_1}(\chi) \right| \cdot \hat{1}_{i, \bar{x}_i, y_i}(\chi)^2 \right)}$$

$$= \sqrt{\prod_{i \in \{2,3\}} \left(\sum_{\chi \neq 1} \sum_{y \in \mathcal{Y}} \mathop{\mathbb{E}}_{x \leftarrow P_X} \left[\left| \hat{1}_{1, \bar{x}_1, y_1}(\chi) \right| \cdot \hat{1}_{i, \bar{x}_i, y_i}(\chi)^2 \right] \right)}.$$

By the independence of (X_1, Y_1) and (X_i, Y_i) under P for $i \in \{2, 3\}$, this is equal to

$$\prod_{i \in \{2,3\}} \sqrt{\sum_{\chi \neq 1} \sum_{y \in \mathcal{Y}} \mathop{\mathbb{E}}_{x \leftarrow P_X} \left[\left| \hat{1}_{1, \bar{x}_1, y_1}(\chi) \right| \right] \cdot \mathop{\mathbb{E}}_{x \leftarrow P_X} \left[\hat{1}_{i, \bar{x}_i, y_i}(\chi)^2 \right]}$$

$$= \prod_{i \in \{2,3\}} \sqrt{\left| \mathcal{Y}_{5-i} \right| \cdot \sum_{\chi \neq 1} \left(\sum_{y_1 \in \mathcal{Y}_1} \mathop{\mathbb{E}}_{x \leftarrow P_X} \left[\left| \hat{1}_{1, \bar{x}_1, y_1}(\chi) \right| \right] \right) \cdot \left(\sum_{y_i \in \mathcal{Y}_i} \mathop{\mathbb{E}}_{x \leftarrow P_X} \left[\hat{1}_{i, \bar{x}_i, y_i}(\chi)^2 \right] \right)}.$$

But the function $1_{1, \bar{x}_1, y_1}$ is just $P_{Y_1 | \bar{X}_1 = \bar{x}_1}(y_1) \cdot \varphi_{1, \bar{x}_1, y_1}$, so the above is

$$\prod_{i \in \{2,3\}} \sqrt{\left| \mathcal{Y}_{5-i} \right| \cdot \sum_{\chi \neq 1} \left(\sum_{y_1 \in \mathcal{Y}_1} \mathop{\mathbb{E}}_{x \leftarrow P_X} \left[P_{Y_1 | \bar{X}_1 = \bar{x}_1}(y_1) \cdot \left| \hat{\varphi}_{1, \bar{x}_1, y_1}(\chi) \right| \right] \right) \cdot \left(\sum_{y_i \in \mathcal{Y}_i} \mathop{\mathbb{E}}_{x \leftarrow P_X} \left[\hat{1}_{i, \bar{x}_i, y_i}(\chi)^2 \right] \right)}$$

$$= \prod_{i \in \{2,3\}} \sqrt{\left| \mathcal{Y}_i \right| \cdot \sum_{\chi \neq 1} \left(\sum_{y_1 \in \mathcal{Y}_1} \mathop{\mathbb{E}}_{x \leftarrow P_X} \left[P_{Y_1 | \bar{X}_1 = \bar{x}_1}(y_1) \cdot \left| \hat{\varphi}_{1, \bar{x}_1, y_1}(\chi) \right| \right] \right) \cdot \left(\sum_{y_i \in \mathcal{Y}_i} \mathop{\mathbb{E}}_{x \leftarrow P_X} \left[\hat{1}_{i, \bar{x}_i, y_i}(\chi)^2 \right] \right)}$$

which by the definition of expectation is

$$\prod_{i \in \{2,3\}} \sqrt{\left| \mathcal{Y}_i \right| \cdot \sum_{\chi \neq 1} \left(\mathop{\mathbb{E}}_{x,y \leftarrow P_{X,Y}} \left[\left| \hat{\varphi}_{1, \bar{x}_1, y_1}(\chi) \right| \right] \right) \cdot \left(\sum_{y_i \in \mathcal{Y}_i} \mathop{\mathbb{E}}_{x \leftarrow P_X} \left[\hat{1}_{i, \bar{x}_i, y_i}(\chi)^2 \right] \right)}.$$

We use Eq. (5) to bound this by

$$\prod_{i\in\{2,3\}} \sqrt{2\epsilon|\mathcal{Y}_i| \cdot \sum_{\chi\neq 1} \sum_{y_i\in\mathcal{Y}_i} \mathop{\mathbb{E}}_{x\leftarrow P_X}\left[\hat{1}_{i,\bar{x}_i,y_i}(\chi)^2\right]}$$

$$\leq \prod_{i\in\{2,3\}} \sqrt{2\epsilon|\mathcal{Y}_i| \cdot \sum_{y_i\in\mathcal{Y}_i} \mathop{\mathbb{E}}_{x\leftarrow P_X}\left[\mathop{\mathbb{E}}_{x'\leftarrow \bar{x}_i}\left[1_{i,\bar{x}_i,y_i}(x')^2\right]\right]} \qquad \text{(Parseval's Theorem)}$$

$$= \prod_{i\in\{2,3\}} \sqrt{2\epsilon|\mathcal{Y}_i| \cdot \mathop{\mathbb{E}}_{x\leftarrow P_X}\left[\mathop{\mathbb{E}}_{x'\leftarrow \bar{x}_i}\left[\sum_{y_i\in\mathcal{Y}_i} 1_{i,\bar{x}_i,y_i}(x')^2\right]\right]}.$$

But for $y_i \neq y_i'$, the supports of $1_{i,\bar{x}_i,y_i}$ and $1_{i,\bar{x}_i,y_i'}$ are disjoint, so this is at most $2\epsilon\sqrt{|\mathcal{Y}_2|\cdot|\mathcal{Y}_3|}$. $\qquad\square$

5 Local Embeddability in Affine Subspaces

In this section we show that the parallel repeated GHZ query distribution has many coordinates in which the GHZ query distribution can be locally embedded, even conditioned on any affine event of low co-dimension. We first recall the notion of a local embedding.

Definition 5.1. *Let Σ be a finite set, let k and n be positive integers, let Q be a probability distribution on Σ^k, and let $\tilde{P}$ be a probability distribution on $\Sigma^{k\times n}$.*

We say that Q is locally embeddable *in the jth coordinate of $\tilde{P}$ if there exists a probability distribution R on a set $\mathcal{R}$ and functions $e_1,\ldots,e_k : \Sigma\times\mathcal{R}\to\Sigma^n$ such that when sampling $q\leftarrow Q$, $r\leftarrow R$, if $\tilde{X}$ denotes the random variable*

$$\tilde{X} \stackrel{\mathsf{def}}{=} \begin{pmatrix} e_1(q_1,r)^\top \\ \vdots \\ e_k(q_k,r)^\top \end{pmatrix},$$

then:

1. *The probability law of $\tilde{X}$ is exactly $\tilde{P}$.*

2. *It holds with probability 1 that $\tilde{X}^j = q$.*

Proposition 5.2. *Let n and m be positive integers with $m < n$. Let Q denote the GHZ query distribution (uniform on the set $\mathcal{Q} = \{x\in\mathbb{F}_2^3 : x_1+x_2+x_3 = 0\}$), and let $\mathcal{W}$ be an affine shift of $\mathcal{V}^3$ for a subspace $\mathcal{V}\leq\mathbb{F}_2^{1\times n}$ of codimension m with $Q^n(\mathcal{W}) > 0$.*

Then there exist at least $n - m$ distinct values of $j\in[n]$ for which Q is locally embeddable in the jth coordinate of $\tilde{P} \stackrel{\mathsf{def}}{=} Q^n|\mathcal{W}$.

Proof. Suppose otherwise. Without loss of generality, suppose that the coordinates that are *not* locally embeddable include the first $n' \stackrel{\mathsf{def}}{=} m+1$ coordinates (otherwise, $\mathcal{V}$ can be permuted to make this so). That is, for each $j\in[n']$, Q is not locally embeddable in the jth coordinate of $\tilde{P}$.

Let the defining equations for $\mathcal{V}$ be written as

$$\mathcal{V} \stackrel{\mathsf{def}}{=} \left\{x\in\mathbb{F}_2^{1\times n} : x\cdot A = 0\right\}$$

for some choice of $A\in\mathbb{F}_2^{n\times m}$, and let $\mathbf{v}\in\mathbb{F}_2^{3\times n}$ be such that $\mathcal{W} = \mathbf{v}+\mathcal{V}^3$.

Because $2^{n'} > 2^m$, the pigeonhole principle implies that there exist two distinct sets $S_0, S_1 \subseteq [n']$ such that

$$\sum_{j\in S_0} A_j = \sum_{j\in S_1} A_j,$$

10

where recall that A_j denotes the jth row of A. Thus, there is a non-empty subset $S \stackrel{\text{def}}{=} S_0 \Delta S_1 \subseteq [n']$ such that

$$\sum_{j \in S} A_j = 0. \tag{7}$$

Fix some such S. We will show that for any $j \in S$, Q is locally embeddable in the jth coordinate of $\tilde{P}$, which is a contradiction. Let X denote the $\mathbb{F}_2^{3 \times n}$-valued random variable given by the identity function.

Claim 5.3. *For any $j \in S$, the distribution $\tilde{P}_{X^j}$ is identical to Q (i.e., uniformly random on Q).*

Proof. Let $j \in S$ be given. It suffices to show that for every $q, q' \in \mathcal{Q}$, there is a bijection $\Phi_{q,q'} : \mathcal{Q}^n \cap \mathcal{W} \to \mathcal{Q}^n \cap \mathcal{W}$ such that $x \in \mathcal{Q}^n \cap \mathcal{W}$ satisfies $x^j = q$ if and only if $y \stackrel{\text{def}}{=} \Phi_{q,q'}(x)$ satisfies $y^j = q'$. Such a bijection $\Phi_{q,q'}$ can be constructed by defining, for all $j' \in [n]$,

$$\Phi_{q,q'}(x)^{j'} = \begin{cases} x^{j'} + q' - q & \text{if } j' \in S \\ x^{j'} & \text{otherwise.} \end{cases}$$

$\Phi_{q,q'}$ clearly is an injective map from $\mathcal{Q}^n$ to $\mathcal{Q}^n$ and satisfies $\Phi_{q,q'}(x)^j = x^j + q' - q$, so the only remaining thing to check is that it indeed maps $\mathcal{W}$ into $\mathcal{W}$. This is true because it preserves $x \cdot A$. Indeed, for any $i \in [3]$,

$$\begin{aligned} \Phi_{q,q'}(x)_i \cdot A &= x_i \cdot A + \sum_{j' \in S} (q'_i - q_i) \cdot A_{j'} \\ &= x_i \cdot A + (q'_i - q_i) \cdot \sum_{j' \in S} A_{j'} \\ &= x_i \cdot A && \text{(by Eq. (7)).} \qquad \square \end{aligned}$$

For any $j \in S$, let $\Delta^{(j)}$ denote the random variable $\left(X^{j'} - X^j\right)_{j' \in S \setminus \{j\}}$.

Claim 5.4. *For any $j \in S$, it holds in $\tilde{P}$ that $\left(\Delta^{(j)}, X^{[n] \setminus S}\right)$ and X^j are independent.*

Proof. Equivalently (using the definition of $\tilde{P}$), let E denote the event that $X \in \mathcal{W}$, i.e. for all $i \in [3]$,

$$(X_i - \mathbf{v}_i) \cdot A = 0.$$

We need to show that in P, the random variables X^j and $\left(\Delta^{(j)}, X^{[n] \setminus S}\right)$ are conditionally independent given E. To show this, we rely on the following fact:

Fact 5.5. *If Y and Z are any independent random variables, and if E is any event that depends only on Z (and occurs with non-zero probability), then Y and Z are conditionally independent given E.*

It is clear that X^j and $\left(\Delta^{(j)}, X^{[n] \setminus S}\right)$ are independent in P. It is also the case that E depends only on $\left(\Delta^{(j)}, X^{[n] \setminus S}\right)$: E is defined by the constraint that for all $i \in [3]$,

$$\begin{aligned} 0 &= (X_i - \mathbf{v}_i) \cdot A \\ &= \sum_{j' \in S} (X_i^{j'} - X_i^j - \mathbf{v}_i^{j'}) \cdot A_{j'} + \sum_{j' \in [n] \setminus S} (X_i^{j'} - \mathbf{v}_i^{j'}) \cdot A_{j'} && \text{(by Eq. (7))} \\ &= -\mathbf{v}_i^j \cdot A_j + \underbrace{\sum_{j' \in S \setminus \{j\}} (X_i^{j'} - X_i^j - \mathbf{v}_i^{j'}) \cdot A_{j'}}_{\text{depends only on } \Delta^{(j)}} + \underbrace{\sum_{j' \in [n] \setminus S} (X_i^{j'} - \mathbf{v}_i^{j'}) \cdot A_{j'}}_{\text{depends only on } X^{[n] \setminus S}}. && \square \end{aligned}$$

We now put everything togther. Fix any $j \in S$. We construct a local embedding of Q into the jth coordinate of $\tilde{P}$. For each $i \in [3]$, we define $e_i : \mathbb{F}_2 \times (\mathbb{F}_2^{3 \times n}) \to \mathbb{F}_2^{1 \times n}$ such that for each $j' \in [n]$:

$$
e_i(x,r)^{j'} = \begin{cases} x & \text{if } j' = j \\ x + r_i^{j'} - r_i^{j} & \text{if } j' \in S \setminus \{j\} \\ r_i^{j'} & \text{if } j' \notin S. \end{cases}
$$

Define the distribution $P^{(\mathsf{embed})}$ to be the distribution on $x \in \mathbb{F}_2^{3 \times n}$ obtained by independently sampling $q \leftarrow Q$ and $r \leftarrow \tilde{P}$, then defining

$$
x \stackrel{\mathsf{def}}{=} \begin{pmatrix} e_1(q_1, r) \\ e_2(q_2, r) \\ e_3(q_3, r) \end{pmatrix}.
$$

It clearly holds with probability 1 that $q = x^j$.

Claim 5.6. $P^{(\mathsf{embed})} \equiv \tilde{P}$.

Proof. By definition, it is immediate that: $P_{X^j}^{(\mathsf{embed})} \equiv \tilde{P}_{X^j}$ and $P_{\Delta^{(j)}, X^{[n] \setminus S}}^{(\mathsf{embed})} \equiv \tilde{P}_{\Delta^{(j)}, X^{[n] \setminus S}}$.

Finally, X is fully determined by X^j and $(\Delta^{(j)}, X^{[n]\setminus S})$, which are independent in both $P^{(\mathsf{embed})}$ (because q and r are sampled independently in the definition of $P^{(\mathsf{embed})}$) and $\tilde{P}$ (by Claim 5.4). □

We have constructed an embedding of Q into one of the first n' coordinates of $\tilde{P}$, which is the desired contradiction. □

6 Decomposition Into Pseudorandom Affine Components

In this section we show that if E is an arbitrary event with sufficient probability mass under $P = Q_{\mathsf{GHZ}}^n$, then $\tilde{P} = P|E$ can be decomposed into components with affine support that are "similar" to corresponding components of P. We will call such components pseudorandom.

We say that Π is an affine partition of $\mathbb{F}_2^{3 \times n}$ to mean that:

- Each part $\Pi(x)$ of Π has the form $\mathbf{w}(x) + \mathcal{V}(x)^3$ where $\mathcal{V}(x)$ is a subspace of $\mathbb{F}_2^n$, and

- Each $\mathcal{V}(x)$ has the same dimension, which we refer to as the dimension of Π and denote by $\dim(\Pi)$. The codimension of Π is defined to be $n - \dim(\Pi)$.

Definition 6.1. *If $\mathcal{W}$ is an affine shift of a vector space $\mathcal{V}^3$ (for $\mathcal{V} \leq \mathbb{F}_2^n$), we say that a $\mathcal{W}$-valued random variable X is (m, ϵ)-close to Y if for all linear functions $\phi : \mathbb{F}_2^n \to \mathbb{F}_2^m$ we have $d_{\mathsf{KL}}(\phi^3(X) \| \phi^3(Y)) \leq \epsilon$, where ϕ^3 denotes the function mapping*

$$
\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \mapsto \begin{pmatrix} \phi(x_1) \\ \phi(x_2) \\ \phi(x_3) \end{pmatrix}.
$$

We write $d_m(X \| Y)$ to denote the minimum ϵ for which X is (m, ϵ)-close to Y.

We remark that $d_m(X \| Y)$ is a non-decreasing function of m.

Lemma 6.2. *Let P denote the distribution Q_{GHZ}^n, let X be the identity random variable, let E be an event with $P(X \in E) = e^{-\Delta}$, and let $\tilde{P} = P|(X \in E)$. For any $\delta > 0$ and any $m \in \mathbb{Z}^+$, there exists an affine partition Π of $\mathbb{F}_2^{3 \times n}$, of codimension at most $m \cdot \frac{\Delta}{\delta}$, such that:*

$$
\mathbb{E}_{\pi \leftarrow \tilde{P}_{\Pi(X)}} \left[d_m \left(\tilde{P}_{X|X \in \pi} \,\Big\|\, P_{X|X \in \pi} \right) \right] \leq \delta. \tag{8}
$$

Proof. We construct the claimed partition iteratively. Start with the trivial n-dimensional affine partition $\Pi_0 = \{\mathbb{F}_2^{3 \times n}\}$. Whenever Π_i is a partition Π for which Eq. (8) does not hold, there exists a function $\phi_i : \mathbb{F}_2^{3 \times n} \to \mathbb{F}_2^{3 \times m}$ that:

- When restricted to any part π of Π_i, ϕ_i is of the form $\phi_{i,\pi}^3$ for some linear function $\phi_{i,\pi} : \mathbb{F}_2^n \to \mathbb{F}_2^m$, and

-
$$d_{\mathsf{KL}}\left(\tilde{P}_{\phi_i(X)|\Pi_i(X)}\middle\|P_{\phi_i(X)|\Pi_i(X)}\right) > \delta. \tag{9}$$

Without loss of generality, we additionally assume that each $\phi_{i,\pi}$ is "full rank" when restricted to π. That is, if π is an affine shift of $\mathcal{V}^3$, where $\mathcal{V}$ has dimension k, then the restriction of $\phi_{i,\pi}$ to $\mathcal{V}$ is a linear map of rank $\min(k,m)$. It is clear that any $\phi_{i,\pi}$ may be modified to be full rank without decreasing the KL divergence of Eq. (9).

Then by the chain rule for KL divergences,

$$d_{\mathsf{KL}}\left(\tilde{P}_{X|\Pi_i(X),\phi_i(X)}\middle\|P_{X|\Pi_i(X),\phi_i(X)}\right) < d_{\mathsf{KL}}\left(\tilde{P}_{X|\Pi_i(X)}\middle\|P_{X|\Pi_i(X)}\right) - \delta. \tag{10}$$

The left-hand side of Eq. (10) is equivalent to

$$d_{\mathsf{KL}}\left(\tilde{P}_{X|\Pi_{i+1}(X)}\middle\|P_{X|\Pi_{i+1}(X)}\right)$$

with $\Pi_{i+1} = \left\{\pi \cap \{x : \phi_i(x) = z\}\right\}_{\pi \in \Pi_i, z \in \mathbb{F}_2^{3 \times m}}$, which is an affine partition of dimension at least $\dim(\Pi) - m$.

Thus with the non-negative potential function

$$\Phi(\Pi) \stackrel{\mathsf{def}}{=} d_{\mathsf{KL}}\left(\tilde{P}_{X|\Pi(X)}\middle\|P_{X|\Pi(X)}\right),$$

we have $\Phi(\Pi_{i+1}) < \Phi(\Pi_i) - \delta$. But $\Phi(\Pi_0) = -\ln(P(E)) = \Delta$, so there must exist $i^\star \leq \frac{\Delta}{\delta}$ for which Eq. (8) holds with $\Pi = \Pi_{i^\star}$, which has co-dimension at most $m \cdot \frac{\Delta}{\delta}$. $\square$

7 Pseudorandomness Preserves Hardness

Proposition 7.1. *Let $\mathcal{W} \subseteq \mathbb{F}_2^{3 \times n}$ be an affine shift of a linear subspace $\mathcal{V}^3$ and let P be a the uniform distribution on $\{w \in \mathcal{W} : w_1 + w_2 + w_3 = 0\}$, which we assume to be non-empty. Let X denote the identity random variable, let $E = E_1 \times E_2 \times E_3$ be an event with $P(X \in E) = e^{-\Delta}$, and define $\tilde{P} \stackrel{\mathsf{def}}{=} P|(X \in E)$. Suppose that $\tilde{P}_X$ is $(\lceil\frac{1}{\delta}\rceil, \delta)$-close to P_X as in Definition 6.1, for δ satisfying $\delta \leq \min(\frac{\Delta^2}{32} \cdot e^{-4\Delta/\epsilon}, \frac{\Delta^2}{32e^2}, 2\epsilon^2)$.*
Then for each $j \in [n]$, we have $v^j(\mathcal{G}_{\mathsf{GHZ}}^n|\tilde{P}) \leq v^j(\mathcal{G}_{\mathsf{GHZ}}^n|P) + 2\epsilon$.

Proof. Fix $j \in [n]$ to be any coordinate, and let $\tilde{f} = \tilde{f}_1 \times \tilde{f}_2 \times \tilde{f}_3 : \mathcal{W} \to \mathbb{F}_2^3$ be an arbitrary strategy. Let Y denote $\tilde{f}(X)$.

Claim 7.2. *There exists a subspace $\mathcal{U} \leq \mathcal{V}$ of codimension at most $\lceil\frac{1}{\delta}\rceil$ such that:*

- *The jth coordinate x^j of $x \in \mathbb{F}_2^{3 \times n}$ depends only on $x + \mathcal{U}^3$.*

- *For all $\chi \in \hat{\mathcal{U}}$,*
$$\mathbb{E}_{(x,y) \leftarrow P_{X,Y}}\left[d_{\mathsf{KL}}\left(P_{\chi(X_1)|X \in x + \mathcal{U}^3, Y_1 = y_1}\middle\|U_{\chi(X_1)|X \in x + \mathcal{U}^3}\right)\right] \leq \delta,$$

where U denotes the uniform distribution on $\mathcal{W}$.

13

Proof. Start with $\mathcal{U}_1 = \{u \in \mathcal{V} : u^j = 0\}$ (this ensures that any subspace $\mathcal{U} \le \mathcal{U}_1$ satisfies the first desired property). Define a potential function

$$Z(\mathcal{U}) \stackrel{\mathsf{def}}{=} \dim(\mathcal{U}) - \mathop{\mathbb{E}}_{(x,y)\leftarrow P_{X,Y}} \left[H(X_1 | X_1 \in x_1 + \mathcal{U}, Y_1 = y_1) \right],$$

which is clearly non-negative. Additionally, $Z(\mathcal{U})$ (and in particular $Z(\mathcal{U}_1)$) is at most 1 because for any subspace $\mathcal{U} \le \mathcal{V}$ and any $x_1 \in \mathcal{V}$, the entropy chain rule implies

$$\mathop{\mathbb{E}}_{y \leftarrow P_{Y|X_1 \in x_1 + \mathcal{U}}} \left[H(X_1 | X_1 \in x_1 + \mathcal{U}, Y_1 = y_1) \right] = H(X_1 | X_1 \in x_1 + \mathcal{U}) - H(Y_1 | X_1 \in x_1 + \mathcal{U})$$

$$\ge \dim(\mathcal{U}) - 1.$$

(in the first step we used the fact that Y_1 is a function of X_1.

For $i \ge 1$, define $\chi_i \in \hat{\mathcal{U}}_i \setminus \{1\}$ to maximize

$$b_i \stackrel{\mathsf{def}}{=} \mathop{\mathbb{E}}_{(x,y)\leftarrow P_{X,Y}} \left[d_{\mathsf{KL}} \left(P_{\chi_i(X_1)|X \in x + \mathcal{U}_i^3, Y_1 = y_1} \,\middle\|\, U_{\chi_i(X_1)|X \in x + \mathcal{U}_i^3} \right) \right]$$

$$= \mathop{\mathbb{E}}_{(x,y)\leftarrow P_{X,Y}} \left[d_{\mathsf{KL}} \left(P_{\chi_i(X_1)|X \in x + \mathcal{U}_i^3, Y_1 = y_1} \,\middle\|\, \mathrm{Unif}_{\{\pm 1\}} \right) \right]$$

$$= \mathop{\mathbb{E}}_{(x,y)\leftarrow P_{X,Y}} \left[d_{\mathsf{KL}} \left(P_{\chi_i(X_1)|X_1 \in x_1 + \mathcal{U}_i, Y_1 = y_1} \,\middle\|\, \mathrm{Unif}_{\{\pm 1\}} \right) \right]$$

$$= 1 - \mathop{\mathbb{E}}_{(x,y)\leftarrow P_{X,Y}} \left[H\big(\chi_i(X_1) | X_1 \in x_1 + \mathcal{U}_i, Y_1 = y_1 \big) \right],$$

and define $\mathcal{U}_{i+1} \stackrel{\mathsf{def}}{=} \{u \in \mathcal{U}_i : \chi_i(u) = 1\}$. By the entropy chain rule, we have $Z(\mathcal{U}_{i+1}) \le Z(\mathcal{U}_i) - b_i$.

Since the initial potential is at most 1, and all potentials are at least 0, there must be some $i^\star \le \lceil \frac{1}{\delta} \rceil$ for which $b_{i^\star} \le \delta$. The corresponding $\mathcal{U}_{i^\star}$ is the desired subspace of $\mathcal{V}$. $\qquad\square$

Now let $\mathcal{U}$ be as given by Claim 7.2. By Lemma 4.4, we have

$$\mathop{\mathbb{E}}_{x \leftarrow P_X} \left[d_{\mathsf{TV}} \left(P_{Y|X \in x + \mathcal{U}^3}, \prod_{i \in [3]} P_{Y_i | X_i \in x_i + \mathcal{U}} \right) \right] \le \sqrt{2\delta}.$$

By assumption of Proposition 7.1 (together with Pinsker's inequality), $P_{X+\mathcal{U}^3}$ and $\tilde{P}_{X+\mathcal{U}^3}$ are $\sqrt{\frac{\delta}{2}}$-close in total variational distance. We thus have that

$$\mathop{\mathbb{E}}_{x \leftarrow \tilde{P}_X} \left[d_{\mathsf{TV}} \left(P_{Y|X \in x + \mathcal{U}^3}, \prod_{i \in [3]} P_{Y_i | X_i \in x_i + \mathcal{U}} \right) \right] \le \sqrt{8\delta}, \tag{11}$$

by the general fact that if P and Q are two distributions that are ϵ-close in total variational distance, and if X is a B-bounded random variable, then $\big| \mathbb{E}_P[X] - \mathbb{E}_Q[X] \big| \le 2B\epsilon$.

We now obtain a probabilistic lower bound on $P(E|X + \mathcal{U}^3)$. We first lower bound its log-expectation:

$$\mathop{\mathbb{E}}_{x \leftarrow \tilde{P}_X} \left[-\ln P\big(E|X \in x + \mathcal{U}^3\big) \right] = \mathop{\mathbb{E}}_{x \leftarrow \tilde{P}_X} \left[d_{\mathsf{KL}}\big(\tilde{P}_{X|X \in x + \mathcal{U}^3} \| P_{X|X \in x + \mathcal{U}^3} \big) \right]$$

$$\le d_{\mathsf{KL}}\big(\tilde{P}_X \| P_X \big) \qquad\qquad \text{(Fact A.17)}$$

$$\le \Delta.$$

Markov's inequality then implies that for any τ,

$$\mathop{\mathrm{Pr}}_{x \leftarrow \tilde{P}_X} \left[P(E|X \in x + \mathcal{U}^3) \le \tau \right] \le \frac{\Delta}{\ln(1/\tau)}. \tag{12}$$

14

Combining Eq. (12) with Eq. (11) and Fact A.18, we get

$$\mathop{\mathbb{E}}_{x \leftarrow \tilde{P}_X} \left[d_{\mathsf{TV}}\big(\tilde{P}_{Y|X \in x + \mathcal{U}^3}, \prod_{i \in [3]} (P|X_i \in E_i)_{Y_i|X_i \in x_i + \mathcal{U}}\big) \right] \le \frac{\Delta}{\ln(1/\tau)} + \frac{4\sqrt{2\delta}}{\tau}.$$

Since this holds for all $\tau \in [0, 1]$ and because $\delta \le \frac{\Delta^2}{32e^2}$, Corollary C.2 implies that

$$\mathop{\mathbb{E}}_{x \leftarrow \tilde{P}_X} \left[d_{\mathsf{TV}}\big(\tilde{P}_{Y|X \in x + \mathcal{U}^3}, \prod_{i \in [3]} (P|X_i \in E_i)_{Y_i|X_i \in x_i + \mathcal{U}}\big) \right] \le \frac{4\Delta}{\ln\left(\frac{\Delta}{\sqrt{32\delta}}\right)} \le \epsilon, \tag{13}$$

where the last inequality follows from our assumption that $\delta \le \frac{\Delta^2}{32} \cdot e^{-4\Delta/\epsilon}$.

Putting everything together, we have

$$\tilde{P}_{X + \mathcal{U}^3, Y} = \tilde{P}_{X + \mathcal{U}^3} \tilde{P}_{Y|X + \mathcal{U}^3}$$

$$\approx_\epsilon \tilde{P}_{X + \mathcal{U}^3} \cdot \prod_{i \in [3]} (P|X_i \in E_i)_{Y_i|X_i + \mathcal{U}}$$

$$\approx_{\sqrt{\frac{\delta}{2}}} P_{X + \mathcal{U}^3} \cdot \prod_{i \in [3]} (P|X_i \in E_i)_{Y_i|X_i + \mathcal{U}},$$

where $\approx$ denotes closeness in total variational distance.

But $P_{X + \mathcal{U}^3} \cdot \prod_{i \in [3]} (P|X_i \in E_i)_{Y_i|X_i + \mathcal{U}}$ is just the distribution on $(x + \mathcal{U}^3, y)$ obtained by sampling $x \leftarrow P_X$, $y \leftarrow F(x)$, where $F = F_1 \times F_2 \times F_3$ is the following randomized strategy. On input x_i, F_i uses local randomness to sample and output $y_i \leftarrow (P|X_i \in E_i)_{Y_i|X_i \in x_i + \mathcal{U}}$. By Fact 3.8, the probability that $W(x^j, y) = 1$ (which is well-defined because x^j is a function of $x + \mathcal{U}^3$) is at most $v^j(\mathcal{G}_{\mathsf{GHZ}}^n|P)$.

We thus have

$$v^j[\tilde{f}](\mathcal{G}_{\mathsf{GHZ}}^n|\tilde{P}) = \tilde{P}_{X + \mathcal{U}^3, Y}\big(W(X^j, Y) = 1\big)$$

$$\le v^j(\mathcal{G}_{\mathsf{GHZ}}^n|P) + \epsilon + \sqrt{\frac{\delta}{2}}$$

$$\le v^j(\mathcal{G}_{\mathsf{GHZ}}^n|P) + 2\epsilon.$$

Since this holds for arbitrary $\tilde{f}$, we have $v^j(\mathcal{G}_{\mathsf{GHZ}}^n|\tilde{P}) \le v^j(\mathcal{G}_{\mathsf{GHZ}}^n|P) + 2\epsilon$. $\square$

8 Proof of Main Theorem

Theorem 8.1. *If $\mathcal{G} = (\mathcal{X}, \mathcal{Y}, Q, W)$ denotes the GHZ game, then $v(\mathcal{G}^n) \le n^{-\Omega(1)}$.*

Proof. Recall $v(\mathcal{G}) = 3/4$.

Let P denote Q^n; that is P is uniform on $\big\{(X_1, X_2, X_3) \in \mathbb{F}_2^{3 \times n} : X_1 + X_2 + X_3 = 0\big\}$. Let $E = E_1 \times E_2 \times E_3$ be any product event in $\mathbb{F}_2^{3 \times n}$ with $P(E) \ge e^{-\Delta}$ (where Δ is a parameter we will specify later), and let $\tilde{P}$ denote $P|E$.

Let $\delta > 0$ be a parameter we will specify later, and let $m = \lceil \frac{1}{\delta} \rceil$. Recall our definition of d_m (Definition 6.1). Lemma 6.2 states that there exists an affine partition Π of $\mathbb{F}_2^{3 \times n}$, of codimension at most $m \cdot \frac{\Delta}{\delta}$, such that:

$$\mathop{\mathbb{E}}_{\pi \leftarrow \tilde{P}_{\Pi(X)}} \left[d_m\big(\tilde{P}_{X|X \in \pi} \big\| P_{X|X \in \pi}\big) \right] \le \delta.$$

15

Moreover,

$$\mathop{\mathbb{E}}_{\pi \leftarrow \tilde{P}_{\Pi(X)}}\left[d_\infty\left(\tilde{P}_{X|X\in\pi}\middle\|P_{X|X\in\pi}\right)\right] = d_{\mathsf{KL}}\left(\tilde{P}_{X|\Pi(X)}\middle\|P_{X|\Pi(X)}\right)$$

$$\leq d_{\mathsf{KL}}\left(\tilde{P}_X\|P_X\right)$$

$$\leq \Delta.$$

Markov's inequality thus implies that with probability at least $1/3$ when sampling $\pi \leftarrow \tilde{P}_{\Pi(X)}$, it holds that $d_m\left(\tilde{P}_{X|X\in\pi}\middle\|P_{X|X\in\pi}\right) \leq 3\delta$ and $d_\infty\left(\tilde{P}_{X|X\in\pi}\middle\|P_{X|X\in\pi}\right) \leq 3\Delta$. Call such a π pseudorandom, and let $\mathcal{R}$ denote the set of pseudorandom π.

By Proposition 7.1, for each pseudorandom π we have

$$v^j\left(\mathcal{G}^n|(\tilde{P}|\pi)\right) \leq v^j\left(\mathcal{G}^n|(P|\pi)\right) + 2\epsilon \tag{14}$$

as long as

$$3\delta \leq \min\left(\frac{9\Delta^2}{32} \cdot e^{-12\Delta/\epsilon}, \ \frac{9\Delta^2}{32e^2}, \ 2\epsilon^2\right), \tag{15}$$

where ϵ is a parameter we will specify later.

By Proposition 5.2, for each $\pi \in \Pi$ (with $P(\pi) > 0$), it holds for all but $m \cdot \frac{\Delta}{\delta}$ values of $j \in [n]$, we have $v^j\left(\mathcal{G}^n|(P|\pi)\right) = v(\mathcal{G}) = 3/4$. By averaging, there exists some $j^\star \in [n]$ such that

$$\mathop{\mathbb{E}}_{\pi \leftarrow \tilde{P}_{\Pi(X)|\Pi(X)\in\mathcal{R}}}\left[v^{j^\star}\left(\mathcal{G}^n|(P|\pi)\right)\right] \leq \frac{m\Delta}{n\delta} + \left(1 - \frac{m\Delta}{n\delta}\right) \cdot \frac{3}{4},$$

which is at most $7/8$ if

$$\delta \geq \frac{2m\Delta}{n}. \tag{16}$$

Putting everything together, we have

$$v^{j^\star}\left(\mathcal{G}^n|\tilde{P}\right) \leq \mathop{\mathbb{E}}_{\pi \leftarrow \tilde{P}_{\Pi(X)}}\left[v^{j^\star}\left(\mathcal{G}^n|(\tilde{P}|\pi)\right)\right]$$

$$\leq \mathop{\Pr}_{\pi \leftarrow \tilde{P}_{\Pi(X)}}[\pi \notin \mathcal{R}] + \mathop{\Pr}_{\pi \leftarrow \tilde{P}_{\Pi(X)}}[\pi \in \mathcal{R}] \cdot \mathop{\mathbb{E}}_{\pi \leftarrow \tilde{P}_{\Pi(X)|\Pi(X)\in\mathcal{R}}}\left[v^{j^\star}\left(\mathcal{G}^n|(\tilde{P}|\pi)\right)\right]$$

$$\leq \frac{2}{3} + \frac{1}{3} \cdot \left(\frac{7}{8} + 2\epsilon\right)$$

$$\leq \frac{47}{48}$$

if Eqs. (15) and (16) are satisfied and if $\epsilon \leq \frac{1}{32}$. Setting $\epsilon = \frac{1}{32}$, $\Delta = 0.0005 \ln n$, $\delta = n^{-0.4}$, $m = n^{0.4}$ ensures that these constraints are all satisfied for sufficiently large n.

Applying Lemma 8.2 below with $\rho(n) = n^{-0.0005}$ and $\epsilon = \frac{1}{48}$ completes the proof. $\qquad\square$

Lemma 8.2 (Parallel Repetition Criterion). *Let $\mathcal{G} = (\mathcal{X}, \mathcal{Y}, Q, W)$ be a game, and let P denote Q^n. Suppose $\rho : \mathbb{Z}^+ \to \mathbb{R}$ is a function with $\rho(n) \geq e^{-O(n)}$ and $\epsilon > 0$ is a constant such that for all $E = E_1 \times \cdots E_k \subseteq \mathcal{X}^n$ with $P^n(E) \geq \rho(n)$ there exists j such that $v^j\left(\mathcal{G}^n|(P|E)\right) \leq 1 - \epsilon$. Then*

$$v(\mathcal{G}^n) \leq \rho(n)^{\Omega(1)}.$$

Proof. Fix any $f = f_1 \times \cdots \times f_k : \mathcal{X}^n \to \mathcal{Y}^n$. Consider the probability space defined by sampling $X \leftarrow P^n$, and let $Y = f(X)$. We define additional random variables $J_1, \ldots, J_n \in [n]$ and $Z_1, \ldots, Z_n \in \mathcal{X} \times \mathcal{Y}$ where J_1 is an arbitrary fixed value, $Z_i \overset{\mathsf{def}}{=} (X^{J_i}, Y^{J_i})$ for all i, and J_{i+1} depends deterministically on $Z_{\leq i} \overset{\mathsf{def}}{=} (Z_1, \ldots, Z_i)$ as

follows. When $Z_{\leq i} = z_{\leq i}$, J_{i+1} is defined to be a value $j \in [n]$ that minimizes $P^n\big(W(X^j, Y^j) = 1 \big| Z_{\leq i} = z_{\leq i}\big)$. With these definitions, each event $\{Z_{\leq i} = z_{\leq i}\}$ is a product event. In particular, if $P^n(Z_{\leq i} = z_{\leq i}) \geq \rho(n)$ then $P^n\big(W(X^{J_{i+1}}, Y^{J_{i+1}}) = 1 \big| Z_{\leq i} = z_{\leq i}\big) \leq 1 - \epsilon$.

Let WIN_i denote the event that $W(Z_i) = 1$, let $\mathrm{WIN}_{\leq i}$ denote the event $\mathrm{WIN}_1 \wedge \cdots \wedge \mathrm{WIN}_i$, and let w_i denote $P^n\big(\mathrm{WIN}_{\leq i}\big)$. Since $\mathrm{WIN}_{\leq i}$ is the union of some subset of the $|\mathcal{X}|^i \cdot |\mathcal{Y}|^i$ disjoint product events $\{Z_{\leq i} = z_{\leq i}\}$, we have

$$\Pr_{z_{\leq i} \leftarrow P^n_{Z_{\leq i}|\mathrm{WIN}_{\leq i}}} \left[P^n(Z_{\leq i} = z_{\leq i}) \geq \rho(n) \right] \geq 1 - |\mathcal{X}|^i \cdot |\mathcal{Y}|^i \cdot \frac{\rho(n)}{w_i}.$$

Moreover, for all $z_{\leq i}$ for which $P^n(Z_{\leq i} = z_{\leq i}) \geq \rho(n)$, we know that $P^n\big(\mathrm{WIN}_{i+1} \big| Z_{\leq i} = z_{\leq i}\big) \leq 1 - \epsilon$. Thus as long as $w_i \geq 2 \cdot |\mathcal{X}|^i \cdot |\mathcal{Y}|^i \cdot \rho(n)$, we have

$$
\begin{aligned}
w_{i+1} &= w_i \cdot P^n(\mathrm{WIN}_{i+1} | \mathrm{WIN}_{\leq i}) \\
&= w_i \cdot \mathbb{E}_{z_{\leq i} \leftarrow P^n_{Z_{\leq i}|\mathrm{WIN}_{\leq i}}} \left[P^n(\mathrm{WIN}_{i+1} | Z_{\leq i} = z_{\leq i}) \right] \\
&\leq w_i \cdot \left(\Pr_{z_{\leq i} \leftarrow P^n_{Z_{\leq i}|\mathrm{WIN}_{\leq i}}} \left[P^n\big(Z_{\leq i} = z_{\leq i}\big) < \rho \right] + \Pr_{z_{\leq i} \leftarrow P^n_{Z_{\leq i}|\mathrm{WIN}_{\leq i}}} \left[P^n\big(Z_{\leq i} = z_{\leq i}\big) \geq \rho \right] \cdot (1 - \epsilon) \right) \\
&\leq w_i \cdot \left(\frac{1}{2} + \frac{1}{2} \cdot (1 - \epsilon) \right) \\
&= w_i \cdot \left(1 - \frac{\epsilon}{2} \right)
\end{aligned}
$$

Iterating this inequality as long as the condition $w_i \geq 2 \cdot |\mathcal{X}|^i \cdot |\mathcal{Y}|^i \cdot \rho(n)$ is satisfied, we find $w_{i^\star}$ such that $w_{i^\star} \leq \min\big(2 \cdot |\mathcal{X}|^{i^\star} \cdot |\mathcal{Y}|^{i^\star} \cdot \rho(n), (1 - \frac{\epsilon}{2})^{i^\star}\big)$. This is minimized for $i^\star = \Theta(\log \frac{1}{\rho(n)})$ or $i^\star = n$ and gives $v(\mathcal{G}^n) \leq w_{i^\star} \leq \rho(n)^{\Omega(1)}$. $\qquad \square$

A Probability Theory

We recall the notions of probability theory that we will need.

Definition A.1. *A* probability distribution *on a finite set Ω is a function $P : \Omega \to \mathbb{R}$ satisfying $P(\omega) \geq 0$ for all $\omega \in \Omega$ and $\sum_{\omega \in \Omega} P(\omega) = 1$. We extend the domain of P to 2^Ω by writing $P(E)$ to denote $\sum_{\omega \in E} P(\omega)$ for any "event" $E \subseteq \Omega$.*

Definition A.2. *The* support *of $P : \Omega \to \mathbb{R}$ is the set $\{\omega \in \Omega : P(\omega) > 0\}$.*

Definition A.3. *A Σ-valued* random variable *on a sample space Ω is a function $X : \Omega \to \Sigma$.*

Definition A.4 (Expectations)**.** *If $P : \Omega \to \mathbb{R}$ is a probability distribution and $X : \Omega \to \mathbb{R}$ is a random variable, the* expectation *of X under P, denoted $\mathbb{E}_P[X]$, is defined to be $\sum_{\omega \in \Omega} P(\omega) \cdot X(\omega)$.*

We refer to subsets of Ω as events. We use standard shorthand for denoting events. For instance, if X is a Σ-valued random variable and $x \in \Sigma$, we write $X = x$ to denote the event $\{\omega \in \Omega : X(\omega) = x\}$.

Definition A.5 (Indicator Random Variables)**.** *For any event E, we write 1_E to denote a random variable defined as*

$$1_E(\omega) = \begin{cases} 1 & \text{if } \omega \in E \\ 0 & \text{otherwise.} \end{cases}$$

Definition A.6 (Independence)**.** *Events $E_1, \ldots, E_k \subseteq \Omega$ are said to be* independent *under a probability distribution P if $P(E_1 \cap \cdots \cap E_k) = \prod_{i \in [k]} P(E_i)$. Random variables $X_1, \ldots, X_k$ are said to be independent if the events $X_1 = x_1, \ldots, X_k = x_k$ are independent for any choice of $x_1, \ldots, x_k$.*

Definition A.7 (Conditional Probabilities). *If $P : \Omega \to \mathbb{R}$ is a probability distribution and $E \subseteq \Omega$ is an event with $P(E) > 0$, then the* conditional distribution of P given E *is denoted $(P|E) : \Omega \to \mathbb{R}$ and is defined to be*

$$(P|E)(\omega) = \begin{cases} P(\omega)/P(E) & \text{if } \omega \in E \\ 0 & \text{otherwise.} \end{cases}$$

If X is a random variable and P is a probability distribution, we write P_X to denote the induced distribution of X under P. That is, $P_X(x) = P(X = x)$.

If E is an event, we write $P_{X|E}$ as shorthand for $(P|E)_X$.

Definition A.8 (Entropy). *If $P : \Omega \to \mathbb{R}$ is a probability distribution, the entropy (in nats) of P is*

$$H(P) \stackrel{\text{def}}{=} -\sum_{\omega \in \Omega} P(\omega) \cdot \ln\big(P(\omega)\big).$$

When X is a random variable associated with a probability distribution P, we sometimes write $H(X)$ as shorthand for $H(P_X)$.

Definition A.9 (Conditional Entropy). *If P is a probability measure with random variables X and Y, we write*

$$H(P_{X|Y}) \stackrel{\text{def}}{=} \mathop{\mathbb{E}}_{y \leftarrow P_Y} \left[H(P_{X|Y=y}) \right].$$

Fact A.10 (Chain Rule of Conditional Entropy). *For any probability measure P and any random variables X, Y, it holds that*

$$H(P_{X|Y}) = H(P_{X,Y}) - H(P_Y).$$

A.1   Divergences

Definition A.11 (Total Variational Distance). *If $P, Q : \Omega \to \mathbb{R}$ are two probability distributions, then the* total variational distance between P and Q, *denoted $d_{\mathsf{TV}}(P, Q)$, is*

$$d_{\mathsf{TV}}(P, Q) \stackrel{\text{def}}{=} \max_{E \subseteq \Omega} \left| P(E) - Q(E) \right|.$$

An equivalent definition is

$$d_{\mathsf{TV}}(P, Q) \stackrel{\text{def}}{=} \frac{1}{2} \sum_{\omega \in \Omega} \left| P(\omega) - Q(\omega) \right|$$

Definition A.12 (Kullback-Leibler (KL) Divergence). *If $P, Q : \Omega \to \mathbb{R}$ are probability distributions, the* Kullback-Leibler divergence *of P from Q is*

$$d_{\mathsf{KL}}(P\|Q) \stackrel{\text{def}}{=} \sum_{\omega \in \Omega} P(\omega) \ln\left(\frac{P(\omega)}{Q(\omega)} \right),$$

where terms of the form $p \cdot \ln(p/0)$ are treated as 0 if $p = 0$ and $+\infty$ otherwise, and terms of the form $0 \cdot \ln(0/q)$ are treated as 0.

The following relation between total variational distance and Kullback-Leiber divergence, known as Pinsker's inequality, is of fundamental importance.

Theorem A.13 (Pinsker's Inequality). *For any probability distributions $P, Q : \Omega \to \mathbb{R}$, it holds that $d_{\mathsf{TV}}(P, Q) \leq \sqrt{\frac{1}{2} d_{\mathsf{KL}}(P\|Q)}$.*

Definition A.14 (Conditional KL Divergence)**.** *If $P, Q : \Omega \to \mathbb{R}$ are probability distributions and if W, X, Y, and Z are random variables on Ω, we write*

$$d_{\mathsf{KL}}(P_{W|X} \| Q_{Y|Z}) \overset{\text{def}}{=} \underset{x \leftarrow P_X}{\mathbb{E}} \left[d_{\mathsf{KL}}(P_{W|X=x} \| Q_{Y|Z=x}) \right],$$

which is taken to be $+\infty$ if there exists x with $P_X(x) > 0$ but $Q_Z(x) = 0$.

KL divergence obeys a chain rule analogous to that for entropy.

Fact A.15 (Chain Rule for KL Divergence)**.** *If $P, Q : \Omega \to \mathbb{R}$ are probability distributions and W, X, Y, Z are random variables on Ω, then*

$$d_{\mathsf{KL}}(P_{W,X} \| Q_{Y,Z}) = d_{\mathsf{KL}}(P_X \| Q_Z) + d_{\mathsf{KL}}(P_{W|X} \| P_{Y|Z}).$$

A.2 Conditional KL Divergence

Fact A.16. *If $P : \Omega \to \mathbb{R}$ is a probability distribution and $E \subseteq \Omega$ is an event, then*

$$d_{\mathsf{KL}}\big(P|E \big\| P\big) = \ln\left(\frac{1}{P(E)}\right).$$

Fact A.17. *Let $P, Q : \Omega \to \mathbb{R}$ be probability distributions and let X, Y be random variables on Ω with Y a function of X. Then*

$$d_{\mathsf{KL}}(P_{X|Y} \| Q_{X|Y})] \le d_{\mathsf{KL}}(P_X \| Q_X).$$

Proof. This is well known, but for completeness:

$$\begin{aligned}
d_{\mathsf{KL}}(P_{X|Y} \| Q_{X|Y}) &= d_{\mathsf{KL}}(P_{X,Y} \| Q_{X,Y}) - d_{\mathsf{KL}}(P_Y \| Q_Y) && \text{(chain rule)} \\
&= d_{\mathsf{KL}}(P_X \| Q_X) - d_{\mathsf{KL}}(P_Y \| Q_Y) && (Y \text{ is a function of } X) \\
&\le d_{\mathsf{KL}}(P_X \| Q_X). && \text{(non-negativity of KL)} \qquad \square
\end{aligned}$$

A.3 Conditional Statistical Distance

Fact A.18. *Let $P, Q : \Omega \to \mathbb{R}$ be probability distributions, and let $E \subseteq \Omega$ be an arbitrary event. Then*

$$d_{\mathsf{TV}}(P|E, Q|E) \le \frac{2 \cdot d_{\mathsf{TV}}(P, Q)}{P(E)}.$$

Proof. Suppose for the sake of contradiction that for some $A \subseteq E$, we have

$$|(P|E)(A) - (Q|E)(A)| > \frac{2 d_{\mathsf{TV}}(P, Q)}{P(E)}.$$

Multiplying on both sides by $P(E)$, we obtain

$$|P(A) - P(E) \cdot (Q|E)(A)| > 2 d_{\mathsf{TV}}(P, Q).$$

Since $|P(E) - Q(E)| \le d_{\mathsf{TV}}(P, Q)$ and $(Q|E)(A) \le 1$, we have

$$|P(A) - Q(A)| > d_{\mathsf{TV}}(P, Q),$$

which is a contradiction. $\qquad \square$

Corollary A.19. *Let $P : \Omega \to \mathbb{R}$ be a probability distribution, let X, Y and Z be random variables on Ω, and let E be an event such that $\Pr_{z \leftarrow P_Z}[P(E|Z = z) \ge \delta] \ge 1 - \tau$, and let $\tilde{P}$ denote $P|E$. Then*

$$\underset{z \leftarrow P_Z}{\mathbb{E}} \left[d_{\mathsf{TV}}(\tilde{P}_{X|Z=z}, \tilde{P}_{Y|Z=z}) \right] \le \tau + \frac{2 \cdot \mathbb{E}_{z \leftarrow P_Z} \left[d_{\mathsf{TV}}(P_{X|Z=z}, P_{Y|Z=z}) \right]}{\delta}.$$

Proof.

$$\mathop{\mathbb{E}}_{z \leftarrow P_Z} \left[d_{\mathsf{TV}}(\tilde{P}_{X|Z=z}, \tilde{P}_{Y|Z=z}) \right]$$

$$= \mathop{\mathbb{E}}_{z \leftarrow P_Z} \left[\mathbb{1}_{P(E|Z=z)<\delta} \cdot d_{\mathsf{TV}}(\tilde{P}_{X|Z=z}, \tilde{P}_{Y|Z=z}) + \mathbb{1}_{P(E|Z=z)\geq\delta} \cdot d_{\mathsf{TV}}(\tilde{P}_{X|Z=z}, \tilde{P}_{Y|Z=z}) \right]$$

$$\leq \tau + \mathop{\mathbb{E}}_{z \leftarrow P_Z} \left[\mathbb{1}_{P(E|Z=z)\geq\delta} \cdot d_{\mathsf{TV}}(\tilde{P}_{X|Z=z}, \tilde{P}_{Y|Z=z}) \right]$$

$$\leq \tau + \mathop{\mathbb{E}}_{z \leftarrow P_Z} \left[\mathbb{1}_{P(E|Z=z)\geq\delta} \cdot \frac{2 \cdot d_{\mathsf{TV}}(P_{X|Z=z}, P_{Y|Z=z})}{P(E|Z=z)} \right]$$

$$\leq \tau + \frac{2 \cdot \mathbb{E}_{z \leftarrow P_Z} \left[d_{\mathsf{TV}}(P_{X|Z=z}, P_{Y|Z=z}) \right]}{\delta}. \qquad \square$$

B Fourier Analysis

For any (finite) vector space V over $\mathbb{F}_2$, the character group of V, denoted $\hat{V}$, is the set of group homomorphisms mapping V (viewed as an additive group) to $\{\pm 1\}$ (viewed as a multiplicative group). Each such homomorphism is called a character of V.

We will distinguish the spaces of functions mapping from $V \to \mathbb{R}$ and functions mapping $\hat{V} \to \mathbb{R}$ and view them as two different inner product spaces. For functions mapping $V \to \mathbb{R}$, we define the inner product

$$\langle f, g \rangle \overset{\mathsf{def}}{=} \mathop{\mathbb{E}}_{x \leftarrow V} \left[f(x)g(x) \right],$$

and for functions mapping $\hat{V} \to \mathbb{R}$, we define the inner product

$$\langle \hat{f}, \hat{g} \rangle \overset{\mathsf{def}}{=} \sum_{\chi \in \hat{V}} \hat{f}(\chi) \cdot \hat{g}(\chi).$$

If there is danger of ambiguity, we use $\hat{\langle \cdot, \cdot \rangle}$ to denote the latter inner product, and $\hat{\| \cdot \|}$ to denote its corresponding norm.

Fact B.1. *Given a choice of basis for V, there is a canonical isomorphism between V and $\hat{V}$. Specifically, if $V = \mathbb{F}_2^n$, then the characters of V are the functions of the form*

$$\chi_\gamma(v) = (-1)^{\gamma \cdot v}$$

for $\gamma \in \mathbb{F}_2^n$.

Definition B.2. *For any function $f : V \to \mathbb{R}$, its Fourier transform is the function $\hat{f} : \hat{V} \to \mathbb{R}$ defined by*

$$\hat{f}(\chi) \overset{\mathsf{def}}{=} \langle f, \chi \rangle = \mathop{\mathbb{E}}_{x \leftarrow V} \left[f(x)\chi(x) \right].$$

One can verify that the characters of V are orthonormal. Together with the assumption that V is finite, we can deduce that f is equal to $\sum_{\chi \in \hat{V}} \hat{f}(\chi) \cdot \chi$.

Theorem B.3 (Plancherel)**.** *For any $f, g : V \to \mathbb{R}$,*

$$\langle f, g \rangle = \langle \hat{f}, \hat{g} \rangle.$$

An important special case of Plancherel's theorem is Parseval's theorem:

Theorem B.4 (Parseval)**.** *For any $f : V \to \mathbb{R}$,*

$$\|f\| = \|\hat{f}\|.$$

20

C   Bound on Optimization Problem

Let $W : \mathbb{R}^+ \to \mathbb{R}^+$ denote the inverse of the function $x \mapsto x \cdot e^x$ (W is known in the literature as the (principal branch of the) Lambert W function). We rely on the following theorem:

Theorem C.1 ([HH00, Corollary 2.4]). *There exists a constant C (in particular, $C = \ln\left(1 + \frac{1}{e}\right)$ works) such that for all $y \geq e$,*

$$W(y) \leq \ln y - \ln \ln y + C.$$

The following corollary is more directly suited to our needs.

Corollary C.2. *For any $A, B > 0$ satisfying $A \geq eB$,*

$$\min_{\tau \in (0,1)} \frac{A}{\ln\left(\frac{1}{\tau}\right)} + \frac{B}{\tau} \leq \frac{4A}{\ln(A/B)}.$$

Proof. The minimum is achieved (up to a factor of two) when $\frac{A}{\ln\left(\frac{1}{\tau}\right)} = \frac{B}{\tau}$ because $\frac{A}{\ln\left(\frac{1}{\tau}\right)}$ is monotonically increasing with τ while $\frac{B}{\tau}$ is monotonically decreasing. Making the change of variables $z = -\ln(\tau)$, this is equivalent to $ze^z = \frac{A}{B}$, i.e. $z = W\left(\frac{A}{B}\right)$. This choice of z (or equivalently τ) gives

$$
\begin{aligned}
\frac{A}{\ln\left(\frac{1}{\tau}\right)} + \frac{B}{\tau} &= \frac{2A}{W(A/B)} \\
&= 2B \cdot \frac{A/B}{W(A/B)} \\
&= 2B \cdot \exp\left(W(A/B)\right) && \text{(Definition of } W) \\
&\leq \frac{2A \cdot (1 + e^{-1})}{\ln(A/B)} && \text{(Theorem C.1)} \\
&\leq \frac{4A}{\ln(A/B)}. && \qquad\square
\end{aligned}
$$

References

[BGKW88] Michael Ben-Or, Shafi Goldwasser, Joe Kilian, and Avi Wigderson. Multi-prover interactive proofs: How to remove intractability assumptions. In *STOC*, pages 113–131. ACM, 1988.

[BJKS04] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *J. Comput. Syst. Sci.*, 68(4):702–732, 2004.

[CHTW04] Richard Cleve, Peter Høyer, Benjamin Toner, and John Watrous. Consequences and limits of nonlocal strategies. In *CCC*, pages 236–249. IEEE Computer Society, 2004.

[DHVY17] Irit Dinur, Prahladh Harsha, Rakesh Venkat, and Henry Yuen. Multiplayer parallel repetition for expanding games. In *ITCS*, volume 67 of *LIPIcs*, pages 37:1–37:16. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2017.

[EPR35] Albert Einstein, Boris Podolsky, and Nathan Rosen. Can quantum-mechanical description of physical reality be considered complete? *Physical review letters*, 47(10):777, 1935.

[Fei91] Uriel Feige. On the success probability of the two provers in one-round proof systems. In *Structure in Complexity Theory Conference*, pages 116–123. IEEE Computer Society, 1991.

[FGL$^+$91]   Uriel Feige, Shafi Goldwasser, László Lovász, Shmuel Safra, and Mario Szegedy. Approximating clique is almost NP-complete (preliminary version). In *FOCS*, pages 2–12. IEEE Computer Society, 1991.

[FK91]   H. Furstenberg and Y. Katznelson. A density version of the Hales-Jewett theorem. *Journal d'Analyse Mathématique*, 57(1):64–119, December 1991.

[For89]   Lance Jeremy Fortnow. *Complexity-theoretic aspects of interactive proof systems*. PhD thesis, MIT, 1989.

[FRS94]   Lance Fortnow, John Rompel, and Michael Sipser. On the power of multi-prover interactive protocols. *Theor. Comput. Sci.*, 134(2):545–557, 1994.

[FV96]   Uriel Feige and Oleg Verbitsky. Error reduction by parallel repetition - a negative result. In Steven Homer and Jin-Yi Cai, editors, *CCC*, pages 70–76. IEEE Computer Society, 1996.

[GHZ89]   Daniel M. Greenberger, Michael A. Horne, and Anton Zeilinger. *Going Beyond Bell's Theorem*, pages 69–72. Springer Netherlands, Dordrecht, 1989.

[HH00]   Abdolhossein Hoorfar and Mehdi Hassani. Inequalities on the Lambert W function and hyperpower function. *J. Inequal. Pure and Appl. Math*, 2000.

[Hol09]   Thomas Holenstein. Parallel repetition: Simplification and the no-signaling case. *Theory Comput.*, 5(1):141–172, 2009.

[HY19]   Justin Holmgren and Lisa Yang. The parallel repetition of non-signaling games: counterexamples and dichotomy. In *STOC*, pages 185–192. ACM, 2019.

[MS13]   Carl A. Miller and Yaoyun Shi. Optimal robust self-testing by binary nonlocal XOR games. In *TQC*, volume 22 of *LIPIcs*, pages 254–262. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2013.

[Pol12]   D.H.J. Polymath. A new proof of the density Hales-Jewett theorem. *Annals of Mathematics*, 175(3):1283–1327, May 2012.

[PRW97]   Itzhak Parnafes, Ran Raz, and Avi Wigderson. Direct product results and the GCD problem, in old and new communication models. In Frank Thomson Leighton and Peter W. Shor, editors, *STOC*, pages 363–372. ACM, 1997.

[Raz98]   Ran Raz. A parallel repetition theorem. *SIAM J. Comput.*, 27(3):763–803, 1998.

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

[Ver96]   Oleg Verbitsky. Towards the parallel repetition conjecture. *Theor. Comput. Sci.*, 157(2):277–282, 1996.

[Yue16]   Henry Yuen. A parallel repetition theorem for all entangled games. In *ICALP*, volume 55 of *LIPIcs*, pages 77:1–77:13. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2016.