Electronic Colloquium on Computational Complexity, Report No. 132 (2020)

Towards Stronger Counterexamples to the Log-Approximate-Rank Conjecture

Arkadev Chattopadhyay * Ankit Garg † Suhail Sherif ‡

Abstract

We give improved separations for the query complexity analogue of the log-approximate-rank conjecture i.e. we show that there are a plethora of total Boolean functions on n input bits, each of which has approximate Fourier sparsity at most $O(n^3)$ and randomized parity decision tree complexity $\Theta(n)$. This improves upon the recent work of Chattopadhyay, Mande and Sherif [5] both qualitatively (in terms of designing a large number of examples) and quantitatively (improving the gap from quartic to cubic). We leave open the problem of proving a randomized communication complexity lower bound for XOR compositions of our examples. A linear lower bound would lead to new and improved refutations of the log-approximate-rank conjecture. Moreover, if any of these compositions had even a sub-linear cost randomized communication protocol, it would demonstrate that randomized parity decision tree complexity does not lift to randomized communication complexity in general (with the XOR gadget).

1 Introduction

The Log-Rank Conjecture (LRC) of Lovasz and Saks asserts that two very seemingly different quantities, one the deterministic communication complexity of a total function f (denoted by $D(f)$) and the other the log of the rank of its communication matrix (denoted by (M_f)) over the field of reals, are essentially the same, i.e. within a fixed polynomial of each other. While this thirty year old conjecture remains wide open, it's natural to try upper-bounding the communication complexity of f by *some* function of the rank of M_f. The best such known bound was obtained by Lovett [19], rather recently, which showed that $D(f)$ is at most the square-root of the rank of M_f, ignoring log factors.

A tempting analog of the LRC for randomized communication complexity appears in a book by Lee and Shraibman [18] where it was named as the Log-Approximate-Rank Conjecture (LARC). Informally, this is LRC with deterministic communication complexity replaced by bounded-error randomized complexity of f, and rank replaced by the *approximate rank* of M_f, where the approximation is uniform point-wise. The LARC is important for several reasons. First, it implies the LRC itself [9]. Second, it implies several other central conjectures, like the polynomial equivalence of quantum and classical communication complexity of total functions [3]. Third, every known lower bound, until very recently, was no larger than a small polynomial of the log of the approximate rank. Very recently, Chattopadhyay, Mande and Sherif [5] provided a surprisingly simple counterexample to the LARC, that exponentially separated randomized communication complexity from the log of the approximate rank. In particular, their function f has Alice and Bob holding n bits each, the approximate rank of its $2^n \times 2^n$ communication matrix M_f is merely $O(n^2)$ and yet the randomized communication complexity is $\Theta(\sqrt{n})$.

Some questions immediately arise from the above refutation of the LARC. First, is the refutation optimal? There are two ways to measure optimality. The approximate rank and communication complexity are separated by a 4th power. Is this separation true for all functions i.e. is randomized

*Tata Institute for Fundamental Research, Mumbai. email: arkadev.c@tifr.res.in

†Microsoft Research India, Bengaluru. email: garga@microsoft.com

‡Tata Institute for Fundamental Research, Mumbai. This work was mainly done while the author was at Microsoft Research India, Bengaluru. email: suhail.sherif@gmail.com

ISSN 1433-8092

communication complexity always upper bounded by fourth-root of the approximate rank? Interestingly, Gál and Syed [8] recently showed that quantum communication complexity is upper bounded by at most square-root of the approximate rank but for randomized communication, the best upper bound is still linear in the approximate rank. The second way to view optimality is the extent of the gap achieved between log of the approximate rank and communication complexity. This is $O(\log n)$ vs. $\sqrt{n}$ for the current refutation. Can this gap be widened via other functions? This leads us to, of course, the related problem of finding other counter-examples to LARC. Finding a richer set of counter-examples, besides being interesting in their own right, could prove useful for understanding other central conjectures. A concrete example is the question of relative power of quantum and classical protocols to solve total functions, a major open problem. If we have to find a total function with an exponential gap between the quantum communication and randomized communication complexities (if one exists at all), then the function should also have an exponential separation between log of approximate rank and randomized communication complexity.[1] However, it was shown by Anshu et al. [1] and Sinha and de Wolf [23] that the function of [5] has large quantum communication complexity (hence refuting the quantum version of LARC as well). This motivates the search for other examples refuting the LARC.

In this work, we come up with a rich set of functions that leaves us with the following win-win situation: either every one of these functions gives a stronger refutation of the LARC than what is known or there is no *lifting theorem* for randomized communication complexity of XOR functions. Lifting theorems, in the setting of communication complexity, lift the complexity of a function f in an appropriate query model to the communication complexity of a problem crafted out of f naturally by block composition with a gadget g, denoted by $f \circ g$. Starting with the celebrated work of Raz and McKenzie [22], they have enabled major progress recently in communication complexity and adjoining areas [11, 7, 10, 4]. In all these theorems, the size of the gadget g is at least logarithmic in the input length of the query function f. A challenging open problem is to prove lifting theorems for a constant size gadget. A natural one is the one bit[2] XOR gadget. It is not hard to verify that a (randomized) parity decision tree (R)PDT algorithm for f of cost c readily translates into a communication protocol of cost $2c$ for $f \circ \mathrm{XOR}$. A lifting theorem for XOR functions would assert the converse. In other words, a communication protocol cannot be more efficient than naively simulating the optimal RPDT. The strongest evidence for such an assertion is the result of Hatami, Hosseini and Lovett [16] who show that if f has deterministic PDT cost c, then $f \circ \mathrm{XOR}$ has deterministic communication complexity $c^{\Omega(1)}$. While no general result exists for the randomized model, the community believes it to be plausible. We state our main result informally.

Theorem 1.1 (Informal)**.** *Assuming XOR lifting theorems for randomized communication complexity, there exists a rich class of functions $f : \{0,1\}^n \to \{0,1\}$, such that $M_{f \circ XOR}$ has approximate rank $O(n^3)$ and $R(f \circ XOR) = \Theta(n)$.*

Thus, conditionally, we get the following improvements over the results in [5]: (1) We narrow the gap between approximate rank and randomized communication complexity from quartic to cubic. (2) We expand the gap between log-approximate-rank and randomized complexity from $O(\log n)$ vs. $\sqrt{n}$ to $O(\log n)$ vs. n, thus yielding essentially the strongest possible refutation of the LARC, under plausible assumptions. While this is a nice conceptual way to view our results, it seems proving communication lower bounds for these functions will require new tools and techniques. On the other hand, coming up with non-trivial communication protocol for any of these functions will rule out a PDT to communication lifting theorem for XOR functions in the randomized model.

1.1 Main Ideas

The starting point of our work is to pursue the idea in [5] of looking for functions with small (approximate) spectral norm, i.e. functions whose sum of the magnitude of Fourier coefficients is a small polynomial in n. The previous counterexample to the LARC used the concept of disjoint subcubes

[1] Since log of the approximate rank lower bounds quantum communication as well.

[2] the gadget size here means the number of bits held by each of the two players.

to achieve this as every subcube has spectral norm one. This implied that a function f whose set of ones form a union of polynomially many disjoint subcubes will have polynomial spectral norm. The fact that polynomial spectral norm implies polynomial Fourier sparsity, yields that the approximate rank of every such f lifted by XOR is guaranteed to be small. The randomized communication complexity of one such function, $\mathsf{SINK} \circ \mathsf{XOR}$, was shown to be large via a Corruption Bound, the proof of which utilized Shearer's Lemma. The randomized parity decision tree lower bound used a robust subspace-hitting property of the subcubes instead.

In this work, we study a broader class of functions based on disjoint subspaces. The approximate rank of their lifts by XOR is again guaranteed to be small. The main conceptual contribution of our work is to identify a property that is sufficient for every such union of subspaces to have large RPDT complexity. Remarkably, this property is quite well encapsulated in the concept of Subspace Designs, a notion that has been studied in the literature in the context of error correcting codes and pseudorandomness [14, 13, 15]. We show that Subspace Designs are hard for RPDTs. The general philosophy of LARC like conjectures is that randomized complexity of total functions is well captured/characterized by algebraic or analytical measures of the function like (approximate) rank. For instance, a classical result of Nisan and Szegedy [20] confirms this idea in the world of randomized (and quantum) query complexity where the relevant algebraic measure is approximate degree. In the world of PDTs, the natural algebraic notion is approximate Fourier sparsity. The work of [5] refuted this philosophy for parity decision trees via the SINK function, whose approximate Fourier sparsity is $O(n^2)$ and RPDT complexity is $\Theta(\sqrt{n})$. Our lower bounds for functions based on subspace designs yields unconditionally a stronger refutation of this philosophy for the model of parity decision trees. We state here our result in terms of random subspaces because this yields the cleanest formulation.

Theorem 1.2 (Main Result)**.** *Let $m = 100n$. Let $\mathcal{V} = \{V_1, V_2, \ldots, V_m\}$ be a set of subspaces of $\{0,1\}^n$ chosen independently and uniformly at random from the set of subspaces of dimension $2n/5$. Let f be the function that outputs 1 on the set $\bigcup_{V \in \mathcal{V}} V$. With probability $1 - o(1)$ the following two statements are true.*

- *Randomized parity decision tree complexity of f is at least $\Omega(n)$.*
- *The spectral norm of f (sum of absolute values of its Fourier coefficients) is upper bounded by $O(n)$ and its approximate Fourier sparsity is upper bounded by $O(n^3)$.*

Hence there exist functions which have a merely cubic gap between approximate Fourier sparsity and RPDT complexity.

The two properties of random subspaces appearing in such a collection that we use are the following: each pair of them have no non-trivial intersection. They also form a (dual) subspace design. We are not able to prove non-trivial lower bounds for the communication problems arising out of Subspace Designs composed with the XOR gadget. However, in Section 3.2, we state concrete conjectures, that seem to be interesting from a Fourier analytic and additive combinatorics point of view, which imply linear lower bounds for such communication problems.

1.2 Organization and plan of the paper

Section 2 contains some basic preliminaries. In Section 3, we prove our main result, a lower bound on the RPDT complexity of a natural class of functions arising out of subspace designs. In Section 3.2, we state a few plausible conjectures and show that they imply a lower bound on the communication complexity of functions arising out of subspace designs composed with the XOR gadget. Finally, we end up with some open problems in Section 4.

2 Preliminaries

In this section, we provide some basic preliminaries needed for the paper. Section 2.1 starts off with some notation. Then in Section 2.2, we present some basic facts about subspaces. Then we introduce

the basics of our models of computations, parity decision trees and communication protocols in Section 2.3. Finally, in Section 2.4, we present some basic concepts from Fourier analysis.

2.1 Notation

Given a subspace $S \subseteq \mathbb{F}_2^n$, we use $\mathsf{dim}(S)$ to denote its dimension and $\mathsf{codim}(S)$ to denote its codimension i.e. $n - \mathsf{dim}(S)$. Given the standard bilinear form $\langle \cdot, \cdot \rangle$ on $\mathbb{F}_2^n$, we can define the dual space of S as the set $\{\ell \in \mathbb{F}_2^n \mid \forall x \in S \ \langle \ell, x \rangle = 0\}$. It is a subspace of dimension $n - \mathsf{dim}(S)$ and its dual space is S.

Given a subspace S of dimension k, fix a basis $L = \{\ell_1, \ldots, \ell_{n-k}\}$ of its dual space. For every point $a \in \mathbb{F}_2^{n-k}$, we can define the set $S_a^L = \{x \in \mathbb{F}_2^n \mid \forall i \in [n-k] \ \langle \ell_i, x \rangle = a_i\}$. These are called affine shifts, or cosets, of S. Sets of the kind S_a^L are also called affine subspaces. Each coset of S also has size 2^k. We can also define a coset map of S with respect to a basis of its dual space as

$$\mathsf{coset}_S^L(x) = (\langle \ell_1, x \rangle, \ldots, \langle \ell_{n-k}, x \rangle).$$

It is easy to see that the choice of basis for the dual space does not affect the set of cosets of S. It merely affects the string $a \in \mathbb{F}_2^{n-k}$ that is used to refer to a specific coset. Hence we will refer to the coset map as coset_S, and we may choose an arbitrary basis of the dual space of S in order to interpret the coset map.

From here on, we will use $\{0,1\}$ to refer to $\mathbb{F}_2$. The values 0 and 1 represent the additive and multiplicative identity of $\mathbb{F}_2$.

2.2 Basic facts about subspaces

Here we mention two facts about subspaces that will be useful. We include their proofs in Appendix A.

Lemma 2.1 (Disjoint Subspaces)**.** *Let S be a subspace of $\{0,1\}^n$ of dimension d_1. Let T be a subspace of $\{0,1\}^n$ of dimension d_2 chosen uniformly at random. Then $\Pr_T[S \cap T = \{0\}] \geq 1 - n2^{d_1+d_2-n}$.*

Lemma 2.2. *Let V and W be affine subspaces of $\{0,1\}^n$ satisfying*

$$\frac{|V \cap W|}{|W|} < \frac{|V|}{2^n}.$$

Then $V \cap W = \emptyset$.

2.3 Parity decision trees, communication complexity and the corruption bound

We now define parity decision trees, aimed at computing functions of the form $f : \{0,1\}^n \to \{0,1\}$.

Definition 2.3 (Parity Decision Tree)**.** *A parity decision tree T is a binary tree rooted at a node r satisfying the following properties.*

- *Each internal node is labelled with a set $S \subseteq [n]$.*
- *Each internal node has two children, with one of the edges labelled with a 0 and the other labelled with a 1.*
- *Each leaf has a label from $\{0,1\}$.*

A parity decision tree outputs a value $a \in \{0,1\}$ on given an input $x \in \{0,1\}^n$ as follows. The "current node" below is initialized to the root node r.

- *The tree computes $b = \oplus_{i \in S} x_i$, where S is the label on the current node.*

- *The tree moves to the child that is reached by taking the edge labelled b. If the child is a leaf, output the label of the leaf. Else, repeat the previous step with the child as the current node.*

The cost of the parity decision tree is defined as the height of the tree.

Definition 2.4 (Randomized Parity Decision Tree)**.** *A randomized parity decision tree (RPDT) of cost c is a distribution over deterministic parity decision trees of cost c. The output of the RPDT on an input x is the random variable defined as the output of T on x, where T is a parity decision tree sampled as per the distribution specified by the RPDT.*

The ϵ-error RPDT complexity of a function f, denoted $\mathsf{R}^{\oplus}_{\epsilon}(f)$, is the minimum cost of an RPDT T such that $\forall x, \Pr[f(x) = T(x)] \geq 1 - \epsilon$.

Lemma 2.5 (Corruption, RPDT version)**.** *Let $f : \{0,1\}^n \to \{0,1\}$. Let μ be a distribution on $\{0,1\}^n$ such that $\mu(f^{-1}(0)) = 1/2$. Let $\epsilon \leq 1/8$. Then an ϵ-error cost-c RPDT computing f implies the existence of an affine subspace W such that*

- $\mu(W \cap f^{-1}(1)) \leq 4\epsilon\mu(W)$ *and*
- $\mathsf{codim}(W) \leq c$.

Proof. Note that an ϵ-error cost-c RPDT T computing f implies that for any distribution μ over the inputs of f, there is an RPDT whose expected error, $\mathbb{E}_{T,x\sim\mu}[|T(x) - f(x)|]$, is at most ϵ. Since T is a distribution over deterministic parity decision trees, there is a deterministic parity decision tree whose expected error is also at most ϵ.

Suppose that a subspace such as the one posited in the lemma statement did not exist. Then for any cost-c parity decision tree T, we may compute the error made as follows. Note that the set of inputs that reach any specific leaf forms an affine subspace of codimension at most c, with each pair of such affine subspaces being disjoint. Let $\mathcal{L}$ be the set of these affine subspaces corresponding to the leaves of T that are labelled 0. Then $\sum_{V\in\mathcal{L}} \mu(V) \geq 1/2 - \epsilon$, since otherwise T would be outputting 1 on more than an ϵ mass of 0-inputs. But then $\sum_{V\in\mathcal{L}} \mu(V \cap f^{-1}(1)) \geq \sum_{V\in\mathcal{L}} 4\epsilon\mu(V) \geq 4\epsilon(1/2 - \epsilon) \geq 2\epsilon - 4\epsilon^2 > \epsilon$. So on more than an ϵ mass of 1-inputs, T outputs 0. Hence the tree T is erring on a larger than ϵ mass of inputs and we have a contradiction. □

We now move to communication complexity. We are concerned with the number of bits that two parties Alice and Bob need to communicate in order to compute a function $F : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$. See [17] for a thorough introduction to the topic. We will use that a deterministic communication protocol of cost c partitions the input space of F into at most 2^c rectangles (sets of the form $A \times B$ for $A \subseteq \mathcal{X}, B \subseteq \mathcal{Y}$), and it outputs the same value on all inputs in a rectangle. Randomized communication is defined akin to randomized parity decision trees.

Definition 2.6 (Randomized Communication Protocol)**.** *A randomized communication protocol of cost c is a distribution over deterministic communication protocols of cost c. The output of the randomized communication protocol on an input x is the random variable defined as the output of T on (x, y), where T is a communication protocol sampled as per the distribution specified by the randomized communication protocol.*

The ϵ-error randomized communication complexity of a function F is the minimum cost of an randomized communication protocol T such that $\forall x, y, \Pr[F(x, y) = T(x, y)] \geq 1 - \epsilon$.

The following is a lower-bound technique for randomized communication complexity akin to the lower bound for RPDTs given previously. This technique is well-known with roots in [24].

Lemma 2.7 (Corruption)**.** *Let $F : \{0,1\}^n \to \{0,1\}$. Let ν be a distribution on $\{0,1\}^n$ such that $\nu(F^{-1}(0)) = 1/2$. Let $\epsilon < 1/8$. Then an ϵ-error cost-c randomized communication protocol computing F implies the existence of a rectangle R such that*

- $\nu(R \cap F^{-1}(1)) \leq 4\epsilon\nu(R)$ *and*
- $\nu(R) \geq 2^{-c-3}$.

2.4 Basic notions from Fourier analysis

We now move to Fourier analysis, a particularly useful tool in analyzing Boolean functions. We define the parity functions as follows. For each $S \subseteq [n]$, we define a parity function $\chi_S : \{0,1\}^n \to \{-1,1\}$ as $\chi_S(x) = (-1)^{\sum_{i \in S} x_i}$. These form an orthonormal basis for the class of functions from $\{0,1\}^n$ to $\mathbb{R}$ under the inner product $\langle f, g \rangle = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} f(x)g(x)$. Hence every such function f can be written as $\sum_S \hat{f}(S)\chi_S$. The values $\hat{f}(S)$ are referred to as Fourier coefficients and can be computed as $\langle f, \chi_S \rangle$. Let $\hat{f}$ denote the vector $(\hat{f}(S))_{S \subseteq [n]} \in \mathbb{R}^{2^n}$, known as the Fourier spectrum. We define the following measures of f.

- The sparsity of f is $\|\hat{f}\|_0$.
- The spectral norm of f is $\left\|\widehat{f}\right\|_1$.
- The ϵ-approximate sparsity of f, $\|\hat{f}\|_{0,\epsilon}$, is $\min_{g:\forall x\ |g(x)-f(x)| \leq \epsilon} \|\hat{g}\|_0$.
- The ϵ-approximate spectral norm of f, $\left\|\widehat{f}\right\|_{1,\epsilon}$, is $\min_{g:\forall x\ |g(x)-f(x)| \leq \epsilon} \|\widehat{g}\|_1$.

The Fourier spectrum of a subspace is easy to compute. (See, for instance, [21].) It follows from the spectrum that any subspace $V \subseteq \{0,1\}^n$, the function $\mathbb{1}_V$ satisfies $\left\|\widehat{\mathbb{1}_V}\right\|_1 = 1$.

For a function $f : \{0,1\}^n \to \mathbb{R}$ its composition with XOR, denoted $f \circ \mathsf{XOR}$, is a function $F : \{0,1\}^n \times \{0,1\}^n \to \mathbb{R}$ defined as $F(x,y) = f(x \oplus y)$ where $x \oplus y$ is the bitwise XOR of x and y.

It is a well known fact that for a function $F := f \circ \mathsf{XOR}$, the rank of the communication matrix of F, denoted $\mathsf{rank}(F)$, is equal to $\|\hat{f}\|_0$. The ϵ-approximate rank of F is at most the ϵ-approximate sparsity of f.

We note a theorem useful in showing that a function has small approximate sparsity.

Theorem 2.8 (Grolmusz's Theorem [2, 12, 25, 5])**.** *For any $f : \{0,1\}^n \to \{0,1\}$ and $\delta > \epsilon \geq 0$,*

$$\|\hat{f}\|_{0,\delta} \leq O\left(\left\|\widehat{f}\right\|_{1,\epsilon}^2 n/(\delta - \epsilon)^2\right).$$

We conclude the preliminaries with the useful notion of entropy.

Definition 2.9 (Entropy)**.** *Let X be a discrete random variable. The entropy $H(X)$ is defined as*

$$H(X) := \sum_{s \in \mathrm{supp(X)}} \Pr[X = s] \log\left(\frac{1}{\Pr[X = s]}\right).$$

Fact 2.10 (Folklore)**.** *$|\mathrm{supp}(X)| = k \implies H(X) \leq \log k$, with equality if and only if X is uniform.*

3 The RPDT Complexity of Dual Subspace Designs

In this section, we prove a lower bound on the RPDT complexity of a natural class of functions arising from subspace designs. A subspace design is a set of subspaces such that any small dimensional subspace non-trivially intersects only a few members of the set. (These are referred to as weak subspace designs in [13].)

Definition 3.1 (Subspace Design)**.** *An n-dimensional (s,h)-subspace design is a set of subspaces $\{S_1, S_2, \cdots, S_m\}$ of $\{0,1\}^n$ such that for all subspaces T of dimension at most s, at most h of the m subspaces intersect T non-trivially.*

We call a set of subspaces $\{V_1, V_2, \cdots, V_m\}$ of $\{0,1\}^n$ an n-dimensional (s,h)-dual subspace design if their duals form an (s,h)-subspace design. Dual subspace designs have an alternate characterization based on the notion of independent subspaces.

Definition 3.2 (Independent Subspaces)**.** *Subspaces $S, T \subseteq \{0,1\}^n$ are independent if their coset maps are independent. That is, let L_S and L_T be arbitrary bases for the dual spaces of S and T. For a variable x chosen uniformly at random from $\{0,1\}^n$, consider the random variables $\mathsf{coset}_S(x)$ and $\mathsf{coset}_T(x)$. For every $a \in \mathbb{F}_2^{\mathsf{codim}(S)}, b \in \mathbb{F}_2^{\mathsf{codim}(T)}$, we want that $\Pr[\mathsf{coset}_S(x) = a | \mathsf{coset}_T(x) = b] = \Pr[\mathsf{coset}_S(x) = a] = 2^{-\mathsf{codim}(S)}$.*

In particular this implies that every coset of S intersects with every coset of T.

We now state the alternate characterization of dual subspace designs.

Claim 3.3. *The set $\{V_1, V_2, \cdots, V_m\}$ of $\{0,1\}^n$ is an n-dimensional (s,h)-dual subspace design if and only if for all subspaces W of* codimension *at most s, at least $m-h$ of the m subspaces are* independent *from W.*

This claim follows from the following lemma relating trivial subspace intersections and independent subspaces.

Lemma 3.4 (Independent Subspaces)**.** *Subspaces S and T of $\mathbb{F}_2^n$ are independent if and only if the dual space of S and the dual space of T intersect trivially (i.e. only at the point $0 \in \mathbb{F}_2^n$).*

Proof. Let V and W be the dual spaces of S and T respectively. If V and W intersected at a non-zero point $\ell \in \mathbb{F}_2^n$, then consider bases L_S and L_T for V and W respectively, wherein ℓ is the first element of L_S and also the first element of L_T. The coset maps of S and T with this choice of L_S and L_T cannot be independent since for all $x \in \mathbb{F}_2^n$, the first entries of $\mathsf{coset}_S^{L_S}(x)$ and $\mathsf{coset}_T^{L_T}(x)$ will always agree.

For the other direction, let L_S and L_T be arbitrary bases for V and W respectively. We will show that if V and W intersect trivially, then the coset maps are independent. Assuming V and W intersect trivially, this means that $\mathrm{span}(L_S) \cap \mathrm{span}(L_T) = \{0\}$. Hence $L = L_S \cup L_T$ is an independent set of size $\mathsf{dim}(V) + \mathsf{dim}(W)$. Consider the subspace X with basis L, and let R be its dual subspace. The cosets of R each have size $2^{n-\mathsf{dim}(V)-\mathsf{dim}(W)}$. For any $a \in \mathbb{F}_2^{\mathsf{codim}(S)}, b \in \mathbb{F}_2^{\mathsf{codim}(T)}$, the set $\{x \mid \mathsf{coset}_S^{L_S}(x) = a \wedge \mathsf{coset}_T^{L_T}(x) = b\}$ is a coset of R. Hence $\Pr[\mathsf{coset}_S^{L_S}(x) = a | \mathsf{coset}_T^{L_T}(x) = b] = 2^{-\mathsf{dim}(V)-\mathsf{dim}(W)}/2^{-\mathsf{dim}(W)} = 2^{-\mathsf{dim}(V)}$ □

A useful corollary of Claim 3.3 is that an (s,h)-dual subspace design also forms a hitting set for the set of all affine subspaces of codimension at most s. We will use this fact to lower bound the randomized parity decision tree complexity of unions of subspaces.

Corollary 3.5. *Let $\{V_1, V_2, \cdots, V_m\}$ be an n-dimensional (s,h)-dual subspace design. For all affine subspaces W of* codimension *at most s, at least $m-h$ of the m subspaces intersect with W.*

Proof. This follows from Claim 3.3 and the fact that if two subspaces S and T are independent, then S will intersect any affine shift of T non-trivially. □

We are now ready to prove the main theorem of the section.

Theorem 3.6. *Let $\mathcal{V}$ be an n-dimensional (s,h)-dual subspace design of size m.*

Let f be the function defined as $f^{-1}(1) = \bigcup_{V \in \mathcal{V}} V$. We now show that $\mathsf{R}_\epsilon^{\oplus}(f) \geq s$ as long as $\epsilon < \frac{m-h}{8m} \frac{|f^{-1}(0)|}{2^n}$.

Proof. Consider the distribution μ defined over the inputs of f as follows.

- Sample $z \sim_{\mathsf{unif}} \{0,1\}$.
- If $z = 0$, output a uniformly random input from $f^{-1}(0)$.
- Otherwise, sample $V \sim_{\mathsf{unif}} \mathcal{V}$.
- Output a uniformly random input from V.

Assuming that f is computed by an ϵ-error cost c RPDT, Lemma 2.5 implies the existence of a subspace W such that

- $\mu(W \cap f^{-1}(1)) \leq 4\epsilon\mu(W)$ and
- $\mathsf{codim}(W) \leq c$.

Assume we have a W such that $\mu(W \cap f^{-1}(1)) \leq 4\epsilon\mu(W)$. This means that $\mu(W \cap f^{-1}(1)) \leq \frac{4\epsilon}{1-4\epsilon}\mu(W \cap f^{-1}(0))$. We also know the following from the definition of μ.

$$\mu(W \cap f^{-1}(1)) = 1/2 \cdot \frac{1}{|\mathcal{V}|} \sum_{V \in \mathcal{V}} \frac{|W \cap V|}{|V|}$$

$$\mu(W \cap f^{-1}(0)) = 1/2 \cdot \frac{|W \cap f^{-1}(0)|}{|f^{-1}(0)|} \leq 1/2 \cdot \frac{|W|}{|f^{-1}(0)|}$$

Putting these together, we get that

$$\frac{1}{|\mathcal{V}|} \sum_{V \in \mathcal{V}} \frac{|W \cap V|}{|V|} \leq \frac{4\epsilon}{1-4\epsilon} \frac{|W|}{|f^{-1}(0)|}.$$

Now if $\epsilon < \frac{m-h}{8m} \frac{|f^{-1}(0)|}{2^n} \leq \frac{1}{8}$, then $\frac{4\epsilon}{1-4\epsilon} < \frac{m-h}{m} \frac{|f^{-1}(0)|}{2^n}$. This implies that less than $m-h$ subspaces of $\mathcal{V}$ can satisfy $\frac{|W \cap V|}{|V|} \geq \frac{|W|}{2^n}$, and hence more than h of them *must* satisfy $\frac{|W \cap V|}{|V|} < \frac{|W|}{2^n}$. This means that $W \cap V = \emptyset$ (Lemma 2.2). In other words, W is an affine subspace that managed to evade more than h subspaces of $\mathcal{V}$. But by Corollary 3.5, if W is of codimension at most s, then it is disjoint from at most h subspaces of $\mathcal{V}$. So W must be of codimension more than s.

Hence the codimension of W, and thereby the cost of the RPDT, is at least s. □

Remark 3.7. *The above proof would also work for any union of affine subspaces which forms a hitting set for the set of all large affine subspaces the way that the dual subspace design does.*

3.1 Narrowing the gap between RPDT complexity and approximate sparsity to cubic

In this section, we instantiate Theorem 3.6 with random subspaces to get a mere cubic gap between RPDT complexity and approximate sparsity. It is known that there are efficient probabilistic constructions of subspace designs. We go through such a construction here, and use it to show our main theorem.

Theorem 3.8. *Let $m = 100n$. Let $V_1, V_2, \ldots, V_m$ be subspaces of $\{0,1\}^n$ chosen independently and uniformly at random from the set of subspaces of dimension $2n/5$. With probability $1 - o(1)$ the following two statements are true.*

- *$\mathcal{V} = \{V_1, \ldots, V_m\}$ forms an $(n/5, m/10)$-dual subspace design.*
- *Every pair of subspaces in $\mathcal{V}$ intersects trivially.*

Proof. Let W be a fixed affine subspace of $\{0,1\}^n$ of dimension $4n/5$. Let $\mathcal{V} = \{V_1, V_2, \cdots, V_m\}$ be subspaces of $\{0,1\}^n$ chosen independently and uniformly at random from the set of subspaces of dimension $2n/5$.

Since the duals of W and V_1 have dimension $3n/5$ and $n/5$ respectively, the probability that W and V_1 are independent is at least $1 - n2^{-n/5}$ (Lemma 2.1). This is independently true of W and each $V \in \mathcal{V}$. The probability that W is not independent with *at least* $m/10$ of the m subspaces is at most $\binom{m}{m/10}(n2^{-n/5})^{m/10}$.

Since the number of subspaces of dimension $4n/5$ is at most $(2^n)^{4n/5} = 2^{4n^2/5}$, the probability that there exists such a subspace W that is not independent with at least $m/10$ of the subspaces in $\mathcal{V}$ is at most $2^{4n^2/5}\binom{m}{m/10}(n2^{-n/5})^{m/10}$.

Setting $m = 100n$, this upper bound is at most $2^{\cdot 8n^2+100n+10n\log n-2n^2} = o(1)$.

Hence with high probability, $\mathcal{V}$ is an $(n/5, m/10)$-dual subspace design.

Let f be defined as in the theorem statement. Note that since V_1 and V_2 are random subspaces of dimension $2n/5$, the probability that they intersect only at 0 is at least $1 - n2^{-n/5}$. The probability that any two subspaces in $\mathcal{V}$ intersect at more than just 0 is at most $\binom{m}{2} n2^{-n/5} = o(1)$. □

Theorem 1.2 (Main Result)**.** *Let $m = 100n$. Let $\mathcal{V} = \{V_1, V_2, \ldots, V_m\}$ be a set of subspaces of $\{0,1\}^n$ chosen independently and uniformly at random from the set of subspaces of dimension $2n/5$. Let f be the function that outputs 1 on the set $\bigcup_{V\in\mathcal{V}} V$. With probability $1 - o(1)$ the following two statements are true.*

- *Randomized parity decision tree complexity of f is at least $\Omega(n)$.*
- *The spectral norm of f (sum of absolute values of its Fourier coefficients) is upper bounded by $O(n)$ and its approximate Fourier sparsity is upper bounded by $O(n^3)$.*

Hence there exist functions which have a merely cubic gap between approximate Fourier sparsity and RPDT complexity.

Proof. We know from Theorem 3.8 that with probability $1 - o(1)$ the set $\mathcal{V}$ forms an $(n/5, m/10)$-dual subspace design. We also can trivially lower bound $|f^{-1}(0)|/2^n$ by $1 - m2^{-3n/5}$. Since $\mathcal{V}$ is an $(n/5, m/10)$-dual subspace design, we can conclude from Theorem 3.6 that for $\epsilon \le 1/10$, $\mathsf{R}^{\oplus}_{\epsilon}(f) \ge n/5$.

We also know from Theorem 3.8 that with probability $1 - o(1)$, every pair of subspaces from $\mathcal{V}$ intersects trivially. When this event holds, f can be represented as $\sum_{V\in\mathcal{V}} \mathbb{1}_V - (m-1)\mathbb{1}_{V_0}$ where $V_0 = \{0\}$ is the trivial subspace of dimension 0. Since the spectral norm of a subspace is equal to 1, the spectral norm of f is upper bounded by $m + m - 1 < 2m$. Using Theorem 2.8, this also implies that $\|\hat{f}\|_{0,\epsilon} \le O(m^2 n/\epsilon^2) = O(n^3)$ for any constant ϵ.

This concludes the proof of the merely cubic gap. □

3.2 On Extending this to Communication

In this section, we state a plausible conjecture that would imply a lower bound on the randomized communication complexity of XOR compositions of our functions. The proof of this implication is in Appendix B.

In the RPDT lower bound, we showed that in order for an affine subspace to avoid most of the subspaces of a dual subspace design, the codimension of the affine subspace needs to be large. We could hope for a similar statement in the communication world: For a rectangle to put very little mass on most of the subspaces making up a dual subspace design (i.e., puts very little mass on inputs (x, y) such that $x \oplus y$ lies in the subspaces), the mass of the rectangle must be $2^{-\Omega(n)}$. One particularly neat conjecture that would imply that statement is the following, in which $\mathcal{U}_k$ denotes the uniform distribution over k elements.

Conjecture 3.9. *There exist constants $0 < \alpha < 1$, $\beta > 0$ and $k \ge 1$ such that the following holds. Let $\mathcal{V} = \{V_1, \ldots, V_m\}$ be an n-dimensional (s, h)-dual subspace design. Let B_i be the coset map of V_i. Let X be a random variable over $\{0,1\}^n$ such that $\|B_i(X) - \mathcal{U}_{2^{\mathsf{codim}(V_i)}}\|_1 \ge \alpha$ for more than kh values of $i \in [m]$. Then $H(X) \le n - \beta s$.*

The merely cubic gap in the RPDT world used random subspaces. So for extending it to communication, it would be okay for us to bypass dual subspace designs and prove the theorem for random subspaces instead.

Conjecture 3.10. *There exists a constant $0 < \alpha < 1, \beta > 0$ such that the following holds. Let $m = 100n$. Let $V_1, V_2, \ldots, V_m$ be random subspaces of $\{0,1\}^n$ of dimension $2n/5$, and let $B_1, B_2, \cdots, B_m$ be their coset maps. Let X be a random variable over $\{0,1\}^n$ such that $\|B_i(X) - \mathcal{U}_{2^{3n/5}}\|_1 \ge \alpha$ for at least $m/3$ values of $i \in [m]$. Then with high probability, $H(X) \le n - \beta n$.*

First of all note that the conjectures are true when X is the uniform distribution over an affine subspace. To see this, suppose X is the uniform distribution over an affine subspace W. $H(X) \geq n-s$ is the same as saying that $\mathsf{codim}(W) \leq s$. Then by Claim 3.3, for at least $m-h$ of the subspaces $V_1, \ldots, V_m$, V_i and the dual space of W are independent, which implies that $B_i(X)$ will be exactly uniform ($\mathcal{U}_{2^{\mathsf{codim}(V_i)}}$).

We discuss now why the Conjectures 3.9 and 3.10 appear to be a bit tricky to prove. While the conjectures are true for affine subspaces, the number of distributions (or even the number of subsets of $\{0,1\}^n$) are much larger (doubly exponential in n), so the conjectures are a leap of faith in this sense. But we haven't been able to come up with counterexamples and it would be very interesting to do so. The conceptual way to view the conjectures, e.g. Conjecture 3.10 to be concrete, is that if a random variable X has the property that when projected down to $2n/5$ bits in various ways it loses $\Omega(1)$ bits of entropy, then X overall loses $\Omega(n)$ bits of entropy. Shearer's lemma talks about these kind of statements. While in Shearer's lemma, the projections are onto subcubes, there are generalizations called Brascamp-Lieb inequalities which talk about more general projections (e.g. see [6]). However, the Brascamp-Lieb inequalities can at best guarantee an $\Omega(n/k)$-bit entropy loss in X if there is an $\Omega(1)$-bit entropy loss while projecting X to k bits in various ways. What we want is much stronger. This is one difficulty.

The other difficulty is that a Fourier type approach doesn't seem to work either. One can control $\|B_i(X) - \mathcal{U}_{2^{\mathsf{codim}(V_i)}}\|_1$ by bounding the ℓ_2 distance and then trying to bound the Fourier coefficients of the distribution of X on the dual space of V_i. But this doesn't give any meaningful bound (if done in a naive way at least).

We now state the lower bound on the randomized communication complexity of a dual subspace design composed with XOR that we get assuming Conjecture 3.9. For a set of subspaces in n dimensions $\mathcal{V} = \{V_1, V_2, \ldots, V_m\}$, let $f_{\mathcal{V}}$ be the function on n bits that outputs 1 on inputs in $\cup_{V \in \mathcal{V}} V$.

Theorem 3.11. *[Proof in Appendix B] Let us assume Conjecture 3.9 holds with constants α, β and k. Let $\mathcal{V} = \{V_1, V_2, \ldots, V_m\}$ be an n-dimensional (s,h)-dual subspace design and define γ so that $|\cup_{V \in \mathcal{V}} V| = \gamma 2^n$. Let $F = f_{\mathcal{V}} \circ \mathsf{XOR}$. For $\epsilon < \frac{(1-\alpha)^2}{4} \frac{m-2kh}{8m}(1-\gamma)$, the ϵ-error randomized communication complexity of F is at least $\beta s + \log(1-\gamma)$.*

Given this lower bound, we would want to apply it to get a merely cubic gap between randomized communication complexity and approximate rank along the lines of Theorem 3.8.

Corollary 3.12. *Let $\mathcal{V} = \{V_1, V_2, \ldots, V_m\}$ be an $(n/5, m/20k)$-dual subspace design with (1) $m = 200kn$, (2) each subspace having dimension $2n/5$ and (3) every pair of subspaces intersecting trivially. Let $F = f_{\mathcal{V}} \circ \mathsf{XOR}$. Then assuming Conjecture 3.9,*

- *The $1/10$-error randomized communication complexity of F is $\Omega(n)$.*
- *$\mathsf{rank}_{1/10}(F) = O(n^3)$.*

Proof. The size of $F^{-1}(1)$ would be at most $2^n \sum_{V \in \mathcal{V}} |V| \leq 2^{n+2n/5} m = o(2^{2n})$. We can then use Theorem 3.11 to get a lower bound of $\beta n/5$ when $\epsilon < \frac{(1-\alpha)^2}{4} \frac{m-2kh}{8m} \frac{|F^{-1}(0)|}{2^{2n}}$, which is a constant. Since we can use error reduction to go from error $1/10$ to any small constant error with only a constant blow-up in cost, the $1/10$-error randomized communication complexity is also $\Omega(n)$.

The ϵ-approximate rank of $f \circ \mathsf{XOR}$ is known to be at most the ϵ-approximate sparsity of f. As analyzed in Theorem 3.8, $\left\|\widehat{f_{\mathcal{V}}}\right\|_1 \leq 2m$ and $\|\hat{f}_{\mathcal{V}}\|_{0,1/10} \leq O(m^2 n) = O(n^3)$ and hence $\mathsf{rank}_{1/10}(F) \leq O(n^3)$. □

The existence of a dual subspace design as required in the previous corollary follows by changing Theorem 3.8 to set $m = 200kn$. The proof of the modified statement is syntactically identical to the proof of the original statement.

4 Conclusion and open problems

We come up with new and improved refutations of the query complexity analogue of the log-approximate-rank conjecture, following the work of Chattopadhyay, Mande and Sherif [5]. Our examples are derived from subspace designs, a concept which has previously found applications in coding theory and pseudorandomness [14, 13, 15]. A lot of interesting open problems arise from our work, some of which we mention below.

1. **(Communication complexity of XOR composed subspace designs).** What is the randomized communication complexity of dual subspace designs composed with XOR (as studied in Section 3.2)? A lower bound would follow from Conjecture 3.9. If Conjecture 3.9 is false, is there an alternate way to prove the communication lower bound? Since we already have an RPDT lower bound for dual subspace designs, these functions provide a interesting class of functions to study randomized XOR lifting. Currently we cannot even prove that this class of functions do not have large monochromatic rectangles.

2. **(Communication complexity of XOR composed random subspaces).** What is the randomized communication complexity of random subspaces composed with XOR? A lower bound would follow from Conjecture 3.10 which follows from Conjecture 3.9. Even if Conjecture 3.9 is false, Conjecture 3.10 could still be true or perhaps easier to prove. If even Conjecture 3.10 is false, is there an alternate way to prove the communication lower bound, perhaps adapting the technique of [16] to the randomized communication setting? Here also we cannot prove that there are no large monochromatic rectangles.

3. **(Quantum communication complexity of XOR composed subspace designs).** What is the quantum communication complexity of dual subspace designs composed with XOR? Is there a function in this class which has polylogarithmic quantum communication complexity?

4. **(RPDT and approximate sparsity).** What is the optimal gap between RPDT complexity and approximate sparsity? We give examples where the RPDT complexity is at least cube root of the approximate sparsity and also RPDT complexity is easily seen to be at most the approximate sparsity.

References

[1] Anurag Anshu, Naresh Goud Boddu, and Dave Touchette. Quantum log-approximate-rank conjecture is also false. In David Zuckerman, editor, *60th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2019, Baltimore, Maryland, USA, November 9-12, 2019*, pages 982–994. IEEE Computer Society, 2019. doi:10.1109/FOCS.2019.00063.

[2] Jehoshua Bruck and Roman Smolensky. Polynomial threshold functions, AC^0 functions and spectral norms (extended abstract). In *31st Annual Symposium on Foundations of Computer Science, St. Louis, Missouri, USA, October 22-24, 1990, Volume II*, pages 632–641, 1990.

[3] Harry Buhrman and Ronald de Wolf. Communication complexity lower bounds by polynomials. In *Proceedings of the 16th Annual Conference on Computational Complexity*, CCC '01, page 120, USA, 2001. IEEE Computer Society.

[4] Arkadev Chattopadhyay, Michal Koucký, Bruno Loff, and Sagnik Mukhopadhyay. Simulation beats richness: new data-structure lower bounds. In Ilias Diakonikolas, David Kempe, and Monika Henzinger, editors, *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 1013–1020. ACM, 2018.

[5] Arkadev Chattopadhyay, Nikhil S. Mande, and Suhail Sherif. The log-approximate-rank conjecture is false. *J. ACM*, 67(4), June 2020. URL: https://doi.org/10.1145/3396695.

[6] Michael Christ. The optimal constants in Holder-Brascamp-Lieb inequalities for discrete Abelian groups. *arXiv preprint arXiv:1307.8442*, 2013.

[7] Susanna F. de Rezende, Or Meir, Jakob Nordström, Toniann Pitassi, Robert Robere, and Marc Vinyals. Lifting with simple gadgets and applications to circuit and proof complexity. *Electronic Colloquium on Computational Complexity (ECCC)*, 26:186, 2019.

[8] Anna Gál and Ridwan Syed. Upper bounds on communication in terms of approximate rank. *Electronic Colloquium on Computational Complexity (ECCC)*, 26:6, 2019.

[9] Dmitry Gavinsky and Shachar Lovett. En route to the log-rank conjecture: New reductions and equivalent formulations. In *Automata, Languages, and Programming - 41st International Colloquium, ICALP 2014, Copenhagen, Denmark, July 8-11, 2014, Proceedings, Part I*, pages 514–524, 2014.

[10] Mika Göös, Rahul Jain, and Thomas Watson. Extension complexity of independent set polytopes. *SIAM J. Comput.*, 47(1):241–269, 2018. doi:10.1137/16M109884X.

[11] Mika Göös, Toniann Pitassi, and Thomas Watson. Deterministic communication vs. partition number. *SIAM J. Comput.*, 47(6):2435–2450, 2018. doi:10.1137/16M1059369.

[12] Vince Grolmusz. On the power of circuits with gates of low ℓ_1 norms. *Theor. Comput. Sci.*, 188(1-2):117–128, 1997.

[13] Venkatesan Guruswami and Swastik Kopparty. Explicit subspace designs. *Combinatorica*, 36(2):161–185, 2016.

[14] Venkatesan Guruswami and Chaoping Xing. Folded codes from function field towers and improved optimal rate list decoding. In *Proceedings of the Forty-Fourth Annual ACM Symposium on Theory of Computing*, STOC '12, page 339–350, New York, NY, USA, 2012. Association for Computing Machinery.

[15] Venkatesan Guruswami, Chaoping Xing, and Chen Yuan. Subspace designs based on algebraic function fields. In *44th International Colloquium on Automata, Languages, and Programming, ICALP 2017, July 10-14, 2017, Warsaw, Poland*, volume 80 of *LIPIcs*, pages 86:1–86:10. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2017.

[16] Hamed Hatami, Kaave Hosseini, and Shachar Lovett. Structure of protocols for XOR functions. *SIAM J. Comput.*, 47(1):208–217, 2018.

[17] Eyal Kushilevitz and Noam Nisan. *Communication complexity*. Cambridge University Press, 1997.

[18] Troy Lee and Adi Shraibman. Lower bounds in communication complexity. *Foundations and Trends in Theoretical Computer Science*, 3(4):263–398, 2009.

[19] Shachar Lovett. Communication is bounded by root of rank. *J. ACM*, 63(1):1:1–1:9, 2016.

[20] Noam Nisan and Mario Szegedy. On the degree of boolean functions as real polynomials. In *Proceedings of the Twenty-Fourth Annual ACM Symposium on Theory of Computing*, STOC '92, page 462–467, New York, NY, USA, 1992. Association for Computing Machinery. doi:10.1145/129712.129757.

[21] Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014.

[22] R. Raz and P. McKenzie. Separation of the monotone NC hierarchy. In *Proceedings of the 38th Annual Symposium on Foundations of Computer Science*, FOCS '97, page 234, USA, 1997. IEEE Computer Society.

[23] Makrand Sinha and Ronald de Wolf. Exponential separation between quantum communication and logarithm of approximate rank. In David Zuckerman, editor, *60th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2019, Baltimore, Maryland, USA, November 9-12, 2019*, pages 966–981. IEEE Computer Society, 2019. doi:10.1109/FOCS.2019.00062.

[24] Andrew Chi-Chih Yao. Lower bounds by probabilistic arguments (extended abstract). In *24th Annual Symposium on Foundations of Computer Science, Tucson, Arizona, USA, 7-9 November 1983*, pages 420–428. IEEE Computer Society, 1983. doi:10.1109/SFCS.1983.30.

[25] Shengyu Zhang. Efficient quantum protocols for XOR functions. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 1878–1885, 2014.

A Facts About Subspaces

Lemma 2.1 (Disjoint Subspaces)**.** *Let S be a subspace of $\{0,1\}^n$ of dimension d_1. Let T be a subspace of $\{0,1\}^n$ of dimension d_2 chosen uniformly at random. Then* $\Pr_T[S \cap T = \{0\}] \geq 1 - n2^{d_1+d_2-n}$.

Proof. Let us generate T by choosing d_2 vectors $\{v_1, \ldots, v_{d_2}\}$, each vector independent of the previous ones, in order to form a basis for T. The subspace S intersects T trivially if and only if for all $i \in [d_2]$, $v_i \notin \mathrm{span}(\{v_j\}_{j<i} \cup S)$. We call these events $E_1, \ldots, E_{d_2}$. When choosing v_i to add to the basis for T, there are $2^n - 2^{i-1}$ choices, since $|\mathrm{span}(\{v_j\}_{j<i})| = 2^{i-1}$. Conditioned on $E_1, \ldots, E_{i-1}$, we also know that $|\mathrm{span}(\{v_j\}_{j<i} \cup S)| = 2^{i-1+d_1}$. The probability of E_i occurring is

$$\frac{|\left(\{0,1\}^n \setminus \mathrm{span}(\{v_j\}_{j<i})\right) \setminus \mathrm{span}(\{v_j\}_{j<i} \cup S)|}{|\{0,1\}^n \setminus \mathrm{span}(\{v_j\}_{j<i})|} = \frac{|\{0,1\}^n \setminus \mathrm{span}(\{v_j\}_{j<i} \cup S)|}{|\{0,1\}^n \setminus \mathrm{span}(\{v_j\}_{j<i})|}.$$

We can then calculate the probability of $S \cap T = \emptyset$ as

$$\begin{aligned}\Pr\left[\bigcap_{i\in[d_2]} E_i\right] &= \prod_{i=1}^{d_2} \Pr\left[E_i \;\middle|\; E_1, \cdots, E_{i-1}\right] = \prod_{i=1}^{d_2} \frac{2^n - 2^{d_1+i-1}}{2^n - 2^{i-1}} \\ &\geq \left(1 - \frac{2^{d_1+d_2}}{2^n}\right)^{d_2} \geq 1 - \frac{d_2}{2^{n-d_1-d_2}}.\end{aligned}$$

□

Lemma 2.2. *Let V and W be affine subspaces of $\{0,1\}^n$ satisfying*

$$\frac{|V \cap W|}{|W|} < \frac{|V|}{2^n}.$$

Then $V \cap W = \emptyset$.

Proof. Let $\{\langle v_i, x\rangle = a_i\}_{i\in[k]}$ be the constraints defining the affine subspace W. Let $W_0, W_1, \cdots, W_k$ be the affine spaces defined as follows. The constraints for W_j are $\{\langle v_i, x\rangle = a_i\}_{i\in[j]}$. Clearly $W_0 = \{0,1\}^n$ and $W_k = W$.

Now let us assume that $|V \cap W_i| \neq 0$ and is hence an affine subspace. The set $V \cap W_{i+1}$ is the same affine subspace with the added constraint $\langle v_{i+1}, x\rangle = a_{i+1}$.

- If this constraint was already implied by the constraints in $V \cap W_i$, then $|V \cap W_{i+1}| = |V \cap W_i|$.
- If this constraint is incompatible with the constraints in $V \cap W_i$, then $|V \cap W_{i+1}| = 0$.
- If this constraint was independent of the constraints in $V \cap W_i$, then $|V \cap W_{i+1}| = |V \cap W_i|/2$.

Hence $|V \cap W_k|$ is either 0 or is at least $|V \cap W_0|/2^k$. On the other hand, $|W|/2^n = 1/2^k$. Since $V \cap W_k = V \cap W$ and $V \cap W_0 = V$, we can rewrite this as

$$V \cap W \neq \emptyset \implies \frac{|V \cap W|}{|V|} \geq \frac{|W|}{2^n}.$$

□

B Randomized Communication Lower Bound Assuming the Conjecture

In the following lower bound, we assume Conjecture 3.9 to hold with $\alpha = 1/2$. After the proof we discuss how to modify it to hold for other values of α.

Theorem B.1. *Let us assume Conjecture 3.9 holds with $\alpha = 1/2$ and some constants β, k. Let $\mathcal{V} = \{V_1, V_2, \ldots, V_m\}$ be an n-dimensional (s,h)-dual subspace design and define γ so that $|\cup_{V \in \mathcal{V}} V| = \gamma 2^n$. Let $F = f_{\mathcal{V}} \circ \mathsf{XOR}$. For $\epsilon < \frac{m-2kh}{128m}(1-\gamma)$, the ϵ-error randomized communication complexity of F is at least $\beta s + \log(1-\gamma)$.*

Proof. For any $V \in \mathcal{V}$, let $S_V = \{(x,y) \in \{0,1\}^n \times \{0,1\}^n \mid x \oplus y \in V\}$. Note that $|S_V| = 2^n|V|$ and $F^{-1}(1) = \cup_{V \in \mathcal{V}} S_V$. Consider the distribution ν defined over the inputs of F as follows.

- Sample $z \sim_{\mathsf{unif}} \{0,1\}$.
- If $z = 0$, output a uniformly random input from $F^{-1}(0)$.
- Otherwise, sample $V \sim_{\mathsf{unif}} \mathcal{V}$.
- Output a uniformly random input from S_V.

Assuming F is computed by an ϵ-error cost c communication protocol, Lemma 2.7 implies the existence of a rectangle R such that

- $\nu(R \cap F^{-1}(1)) \leq 4\epsilon\nu(R)$ and
- $\nu(R) \geq 2^{-c-3}$.

Assume we have an R such that $\nu(R \cap F^{-1}(1)) \leq 4\epsilon\nu(R)$. This means that $\nu(R \cap F^{-1}(1)) \leq \frac{4\epsilon}{1-4\epsilon}\nu(R \cap F^{-1}(0))$. We also know the following from the definition of ν.

$$\nu(R \cap F^{-1}(1)) = 1/2 \cdot \frac{1}{|\mathcal{V}|} \sum_{V \in \mathcal{V}} \frac{|R \cap S_V|}{|S_V|}$$
$$\nu(R \cap F^{-1}(0)) = 1/2 \cdot \frac{|R \cap F^{-1}(0)|}{|F^{-1}(0)|} \leq 1/2 \cdot \frac{|R|}{|F^{-1}(0)|}$$

Putting these together, we get that

$$\frac{1}{|\mathcal{V}|} \sum_{V \in \mathcal{V}} \frac{|R \cap S_V|}{|S_V|} \leq \frac{4\epsilon}{1-4\epsilon} \frac{|R|}{|F^{-1}(0)|}.$$

Now if $\epsilon < \frac{m-2kh}{128m} \frac{|F^{-1}(0)|}{2^{2n}} < 1/8$, then $\frac{4\epsilon}{1-4\epsilon} < \frac{m-2kh}{16m} \frac{|F^{-1}(0)|}{2^{2n}}$. This implies that less than $m - 2kh$ subspaces of $\mathcal{V}$ can satisfy $\frac{|R \cap S_V|}{|S_V|} \geq \frac{|R|}{16 \cdot 2^{2n}}$, and hence more than $2kh$ of them *must* satisfy $\frac{|R \cap S_V|}{|S_V|} < \frac{|R|}{16 \cdot 2^{2n}}$. Let us fix such a V.

Let coset_V denote the function $\mathsf{coset}_V^{L_V}$ for some fixed basis L_V of the dual space of V. Let $R = A \times B$. Then $\frac{|R \cap S_V|}{|R|}$ is the probability that, when x and y are sampled uniformly at random from

A and B, $\mathsf{coset}_V(x) = \mathsf{coset}_V(y)$. Let A_V be the distribution of $\mathsf{coset}_V(x)$ and B_V be the distribution of $\mathsf{coset}_V(y)$. The condition $\frac{|R \cap S_V|}{|R|} < \frac{|S_V|}{16 \cdot 2^{2n}}$ can be rewritten as

$$\Pr_{x' \sim A_V, y' \sim B_V}[x' = y'] < \frac{|S_V|}{16 \cdot 2^{2n}} = \frac{1}{16 \cdot 2^{\mathsf{codim}(V)}}.$$

It follows that $A_V(S) < 1/4$ where $S = \{y' \mid B_V(y') \geq \frac{1}{4 \cdot 2^{\mathsf{codim}V}}\}$. However, $B_V(S)$ must be at least $3/4$, since $B_V(\overline{S}) \leq 1/4$.

Hence A_V and B_V have total variational distance at least $1/2$, and $\|A_V - B_V\|_1 \geq 1$. By the triangle inequality, $\max\{\|A_V - \mathcal{U}_{2^{\mathsf{codim}(V)}}\|_1, \|B_V - \mathcal{U}_{2^{\mathsf{codim}(V)}}\|_1\} \geq 1/2$.

Hence, either there are more than kh subspaces that satisfy $\|A_V - \mathcal{U}_{2^{\mathsf{codim}(V)}}\| \geq 1/2$ or there are more than kh subspaces that satisfy $\|B_V - \mathcal{U}\| \geq 1/2$. Without loss of generality we assume the former. Now we use our conjecture. The conjecture implies that $H(A) \leq n - \beta s$. Hence $\frac{|R|}{2^{2n}} \leq 2^{-\beta s}$.

We now want to move from $|R|$ being small under the uniform distribution to R being small under ν. We know that $\nu(R \cap F^{-1}(1)) \leq 4\epsilon\nu(R) < \nu(R)/2$, so $\nu(R \cap F^{-1}(0)) \geq \nu(R)/2$. We also know from the definition of ν that

$$\nu(R \cap F^{-1}(0)) = \frac{|R \cap F^{-1}(0)|}{2|F^{-1}(0)|} \leq \frac{|R|}{2 \cdot 2^{2n}} \cdot \frac{2^{2n}}{|F^{-1}(0)|} \leq 2^{-\beta s - 1} \cdot \frac{1}{1-\gamma}.$$

So $\nu(R) \leq 2\nu(R \cap F^{-1}(0)) \leq 2^{-\beta s - 1 - \log(1-\gamma)}$. Hence the cost of the protocol is at least $\beta s + \log(1 - \gamma) - 3$. □

We now explain how to modify the proof assuming the conjecture were true for other values of α. Then the theorem statement would be modified, setting $\epsilon < \frac{(1-\alpha)^2}{4} \frac{m-2kh}{8m}(1-\gamma)$. The proof would go through as it does above, analyzing a rectangle $R = A \times B$.

- We would find more than $2kh$ subspaces V such that $\Pr[A_V = B_V] < \frac{(1-\alpha)^2}{4} \frac{|S_V|}{2^{2n}}$ as is done in the above proof.
- We would then set $S = \{y' \mid B_V(y') \geq \frac{1-\alpha}{2 \cdot 2^{\mathsf{codim}(V)}}\}$. This would mean that $A_V(S) \leq \frac{1-\alpha}{2}$ and $B_V(S) \geq 1 - \frac{1-\alpha}{2}$. Hence $\|A_V - B_V\|_1 \geq 2\alpha$, and one of A or B (wlog, A) satisfies $\|A_V - \mathcal{U}_{2^{\mathsf{codim}(V)}}\|_1 \geq \alpha$ for at least kh subspaces from the dual subspace design.
- The proof would continue as it does above, using the conjecture to conclude that the cost of the protocol would be at least $\beta s + \log(1-\gamma) - 3$, which is $\Omega(s)$ for constant γ.

ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il