Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 124 (2023)

Explicit separations between randomized and deterministic Number-on-Forehead communication

Zander Kelley[*] Shachar Lovett[†] Raghu Meka[‡]

January 3, 2024

Abstract

We study the power of randomness in the Number-on-Forehead (NOF) model in communication complexity. We construct an explicit 3-player function $f : [N]^3 \to \{0,1\}$, such that: (i) there exist a randomized NOF protocol computing it that sends a constant number of bits; but (ii) any deterministic or nondeterministic NOF protocol computing it requires sending about $(\log N)^{1/3}$ many bits. This exponentially improves upon the previously best-known such separation. At the core of our proof is an extension of a recent result of the first and third authors on sets of integers without 3-term arithmetic progressions into a non-arithmetic setting.

1 Introduction

Number-on-Forehead (NOF) communication introduced by Chandra, Furst, and Lipton [CFL83] is a central model in communication complexity. In its basic form, there are three parties, Alice, Bob, and Charlie, who each have an input x, y, z respectively written on their forehead, and their goal is to communicate and compute a function $f(x, y, z)$. The twist, of course, is that Alice only knows the inputs y, z, Bob the inputs x, z, and Charlie the inputs x, y. The main goal is to understand how much communication is needed to compute the function f.

Since its introduction, NOF communication complexity has been extensively studied and is known to have connections to circuit lower bounds, data structure lower bounds, and additive combinatorics [CFL83, Raz00, BPSW06, BPS07, DPV09, LS09, BDPW10, BH12, LLSW16, LPS18, Shr18, AS20, LS21b, LS21a, AB23], among others. In this work, we study the relative power of randomized vs. deterministic and non-deterministic protocols in the NOF model.

Like standard two-party communication, given a function $f : X \times Y \times Z \to \{0,1\}$, one can study NOF communication complexity under several models: deterministic, non-deterministic and randomized protocols. We formally define these later, and note here that they are the natural analogs of the two-party definitions, where for randomized protocols we assume access to public randomness.

[*] Department of Computer Science, University of Illinois at Urbana-Champaign. Supported by NSF CAREER award 2047310. Email: awk2@illinois.edu

[†] Department of Computer Science and Engineering, University of California, San Diego. Supported by NSF DMS award 1953928, NSF CCF award 2006443, and a Simons investigator award. Email: slovett@ucsd.edu

[‡] Department of Computer Science, University of California, Los Angeles. Supported by NSF AF 2007682 and NSF Collaborative Research Award 2217033. Email: raghum@cs.ucla.edu

ISSN 1433-8092

The (complement of) set-disjointness function provides an explicit example of a function which is easy for nondeterministic protocols but is hard for deterministic or randomized protocols, see [CP10] for a survey on the history of this problem. Our focus in this paper is on the other direction; namely explicit functions which are easy for randomized protocols, but are hard for deterministic and non-deterministic protocols.

Despite the rich literature on NOF protocols, the relative power of randomness for NOF protocols is poorly understood. Beame et al. [BDPW10] showed that there are functions $f : [N]^3 \to \{0,1\}$ whose randomized communication complexity is $O(1)$ but whose deterministic (or even non-deterministic) communication complexity is as large as possible, namely $\Omega(\log N)$. However, this separation is existential (based on a counting argument), and no explicit function is known to strongly separate deterministic and randomized NOF protocols. This is in stark contrast to the two-party case, where the equality function[1] has randomized complexity $O(1)$ and deterministic complexity $\Omega(\log N)$.

The best known explicit separation is given by the *Exactly-N* problem, already considered in the seminal work [CFL83] that defined the NOF model: $\mathsf{ExactlyN} : [N]^3 \to \{0,1\}$ is defined by $\mathsf{ExactlyN}(x, y, z) = 1$ if and only if $x + y + z = N$. The randomized communication complexity of $\mathsf{ExactlyN}$ is $O(1)$ while the best known lower bound on its deterministic communication complexity is $\Omega(\log\log N)$ [BGG06,LS21b]. The interest in this function stems from the fact that deterministic lower bounds for it are equivalent to the famous corners problem in additive combinatorics, which asks for the largest subset of $[N]^2$ without a corner[2]. The conjectured bounds in additive combinatorics would imply a lower bound of $\Omega(\sqrt{\log N})$ for the deterministic communication complexity of $\mathsf{ExactlyN}$; but as mentioned, the known lower bounds are exponentially far from it.

Our main result is an explicit construction of a 3-party function $D : [N]^3 \to \{0,1\}$, with constant randomized NOF communication complexity, and with deterministic (and even non-deterministic) NOF communication complexity polynomial in the input length, concretely $\Omega((\log N)^{1/3})$.

Theorem 1.1. *Let q be a prime power and k a large enough constant ($k = 34$ suffices). Let $N = q^k$, and identify $[N]$ with $\mathbb{F}_q^k$. Consider the following 3-player function:*

$$D(x, y, z) = \mathbb{1}[\langle x, y\rangle = \langle x, z\rangle = \langle y, z\rangle],$$

where $\langle \cdot, \cdot \rangle$ denotes the standard inner-product in $\mathbb{F}_q^k$. Then:

1. *The randomized NOF communication complexity of D (with public randomness) is $O(1)$.*
2. *The deterministic or non-deterministic NOF communication complexity of D is $\Omega((\log N)^{1/3})$.*

One of the challenges in showing a result as above is the dearth of techniques for proving lower bounds in the NOF communication complexity. By and large, the main technique for showing lower bounds in the 3-party NOF model is the discrepancy method as introduced in [BNS89]. However, these techniques typically also lower bound the randomized communication complexity. As such, they are not immediately useful for separating deterministic and randomized communication complexity.

Inspired by recent progress on the *three-term arithmetic progressions* problem [KM23], we introduce a new technique for lower bounding deterministic NOF communication complexity. We are

[1] The equality function is $EQ : [N]^2 \to \{0,1\}$, $EQ(x, y) = \mathbb{1}[x = y]$.

[2] A corner is a triple of points $(x, y), (x + h, y), (x, y + h)$.

optimistic that this technique will find potential applications (for instance for the ExactlyN problem), beyond the application in this paper. Below we first give a high-level overview of the main ideas, and then delve into the technical details.

2 Proof overview

We next give a high-level description of the proof of Theorem 1.1 highlighting the new techniques we introduce. The proof involves two modular steps:

- We introduce a notion of *pseudorandomness against cubes* for sets $D \subset [N]^3$. We show that computing membership in any *sparse* pseudorandom set D (i.e., $|D| < N^{3-c}$ for a constant $c > 0$) is hard for deterministic and non-deterministic NOF protocols.
- We then construct sparse pseudorandom sets as above, which are in addition easy for randomized protocols.

The first step above is more intricate, and we describe it first.

2.1 Pseudorandomness against cubes

A cube $C \subset [N]^3$ is a set of the form $C = X \times Y \times Z$ for some subsets $X, Y, Z \subseteq [N]$. At a high level, we say $D \subset [N]^3$ is pseudorandom if for any large enough cube C, the intersection $D \cap C$ *looks random* in two ways:

- Its density when restricted to the cube is comparable to its overall density.
- The marginals of D on the faces of C are close to uniform.

Below we use the following convention: we identify a set $D \subset [N]^3$ with its indicator function $D : [N]^3 \to \{0, 1\}$. The density of a set is $\mathbb{E}[D] = |D|/N^3$.

Definition 2.1 (Pseudorandom set with respect to large cubes)**.** *Let $D \subset [N]^3$ of density $\mu = \mathbb{E}[D]$, and let $\gamma > 0$. Given a cube $C = X \times Y \times Z \subset [N]^3$, we say that D is γ-pseudorandom with respect to C if the following conditions hold:*

1. $\mathbb{E}_{x,y,z \in X \times Y \times Z}[D(x, y, z)] = \mu(1 \pm \gamma)$.
2. $\mathbb{E}_{x,y,z,z' \in X \times Y \times Z \times Z}[D(x, y, z)D(x, y, z')] = \mu^2(1 \pm \gamma)$*, as well as the analogous conditions involving (x, x', y, z) and (x, y, y', z).*

We say that D is γ-pseudorandom with respect to large cubes if D is γ-pseudorandom with respect to any cube C of size $|C| \geq \gamma N^3$.

We show that the above notion of pseudorandomness along with *sparsity* suffices for hardness against non-deterministic NOF protocols (which include as a special case also deterministic NOF protocols).

Theorem 2.2. *Let $D \subset [N]^3$ be a set of size $|D| \leq N^{3-c}$ which is (N^{-c})-pseudorandom with respect to large cubes, for some constant $c > 0$. Then any non-deterministic NOF protocol which computes D must send $\Omega((\log N)^{1/3})$ communication.*

The above result follows as a corollary to a more general result on how pseudorandom sets as defined above interact with *cylinder intersections* which we next describe.

2.2 From pseudorandomness against cubes to cylinder intersections

We use the well-known connection between NOF protocols and *cylinder intersections*.

Definition 2.3 (Cylinder intersection)**.** *A set $T \subset [N]^3$ is a cylinder intersection, if there are sets $S_1, S_2, S_3 \subset [N]^2$ such that*

$$T = \{(x, y, z) : (x, y) \in S_1, (x, z) \in S_2, (y, z) \in S_3\}.$$

Equivalently, using the function notation, we have the indicator function equality

$$T(x, y, z) \equiv S_1(x, y)S_2(x, z)S_3(y, z).$$

We denote $T = CI(S_1, S_2, S_3)$.

It is well-known that if a function D has a non-deterministic NOF protocol which sends b bits, then D can be expressed as the union of 2^b cylinder intersections. It thus suffices to focus on the structure of cylinder intersections, and their relation to pseudorandom sets D. Our main technical theorem in this context shows that any sparse cylinder intersection must also be sparse with respect to a pseudorandom set D.

Theorem 2.4 (Pseudorandom sets for cubes are pseudorandom for cylinder intersections)**.** *Let $D \subset [N]^3$ be γ-pseudorandom with respect to large cubes. Then, D is pseudorandom with respect to cylinder intersections in the following one-sided sense. Let $t \leq \log(1/\gamma)$. For any cylinder intersection F of density*

$$\mathop{\mathbb{E}}_{(x,y,z)\in[N]^3} F(x, y, z) \leq 2^{-t},$$

it holds

$$\mathop{\mathbb{E}}_{(x,y,z)\in D} F(x, y, z) \leq 2^{-ct^{1/3}}$$

for some (absolute) constant $c > 0$.

Theorem 2.4 shows that any set D which is sparse and pseudorandom for large cubes, must be difficult for non-deterministic NOF protocols. Theorem 2.2 follows easily from the above theorem.

Proof of Theorem 2.2 from Theorem 2.4. Assume that a non-deterministic NOF protocol which sends b bits decides $D(x, y, z)$. This implies that D can be expressed as the union of 2^b cylinder intersections $F_1, \ldots, F_{2^b} \subset [N]^3$. For each cylinder intersection F_i, since $F_i \subset D$ and we assume $|D| \leq N^{3-c}$, we have $|F_i| \leq N^{3-c}$. Since we assume that $\gamma = N^{-c}$, we may apply Theorem 2.4 for $t = c \log N$. This implies that for some constant $c' > 0$,

$$|F_i \cap D| \leq 2^{-c'(\log N)^{1/3}}|D|.$$

Since $F_1, \ldots, F_{2^b}$ cover D, their number must satisfy $2^b \geq 2^{c'(\log N)^{1/3}}$, from which the theorem follows. □

We next describe the main ideas in the proof of Theorem 2.4. Its proof relies on a new technical result, that shows that the product of pseudorandom matrices is close to uniform.

2.3 Uniformity from spreadness

We consider matrices with bounded entries. Given a matrix A with rows indexed by X, columns indexed by Y and entries bounded in $[0,1]$, it will be convenient for us to identify it with the function $A : X \times Y \to [0,1]$. We study three properties of matrices each of which measures how close a matrix is to a constant matrix. The first property is *spreadness*, which means that the density of a matrix cannot be significantly increased by restricting to a large rectangle.

Definition 2.5 (Spreadness)**.** *Let $A : X \times Y \to [0,1]$, and let $r \geq 1, \varepsilon \in (0,1)$. We say that A is (r,ε)-spread if for any rectangle $R = X' \times Y' \subset X \times Y$ of size $|R| \geq 2^{-r}|X \times Y|$, it holds that*

$$\mathop{\mathbb{E}}_{(x,y)\in R}[A(x,y)] \leq (1+\varepsilon)\,\mathbb{E}[A].$$

The second property is *left lower-bounded*, which in matrix notation means that row averages are not much lower than the global average of the matrix.

Definition 2.6 (Left lower-bounded)**.** *Let $A : X \times Y \to [0,1]$, and let $\varepsilon \in (0,1)$. We say that A is ε-left lower-bounded if*

$$\mathop{\mathbb{E}}_{y\in Y}[A(x,y)] \geq (1-\varepsilon)\,\mathbb{E}[A] \quad \forall x \in X.$$

The third property is *uniformity*, which means that nearly all the entries of a matrix are multiplicatively close to some fixed value.

Definition 2.7 (Uniformity)**.** *Let $A : X \times Y \to [0,1]$, and let $k \geq 1, \alpha, \varepsilon \in (0,1)$. We say that A is (α,k,ε)-uniform if*

$$\Pr_{(x,y)\in X\times Y}[(1-\varepsilon)\alpha \leq A(x,y) \leq (1+\varepsilon)\alpha] \geq 1-2^{-k}.$$

Given $A : X \times Z \to [0,1]$, $B : Y \times Z \to [0,1]$, define their normalized product $A \circ B : X \times Y \to [0,1]$ as

$$(A \circ B)(x,y) = \mathop{\mathbb{E}}_{z\in Z} A(x,z)B(y,z).$$

If we consider A, B as matrices, then $A \circ B = \frac{1}{|Z|}AB^{\top}$. The following theorem, which can be considered as the main new technical step in the proof, shows that if A, B are both spread and left lower-bounded, then their normalized product $A \circ B$ is uniform, where almost all its elements are close to the value expected in the random case, namely $\mathbb{E}[A]\,\mathbb{E}[B]$.

Theorem 2.8 (Product of spread matrices is uniform)**.** *Let $A : X \times Z \to [0,1], B : Y \times Z \to [0,1]$. Let $d, k \geq 1$ and $\varepsilon \in (0, 1/80)$. Assume that*

1. *$\mathbb{E}[A], \mathbb{E}[B] \geq 2^{-d}$.*
2. *A, B are (r,ε)-spread for $r \geq \Omega(dk/\varepsilon)$.*
3. *A, B are ε-left lower-bounded.*

Then $A \circ B$ is $(\mathbb{E}[A]\,\mathbb{E}[B], k, 80\varepsilon)$ uniform.

We note that a corollary of Theorem 2.8 is that $\mathbb{E}[A \circ B] \approx \mathbb{E}[A]\,\mathbb{E}[B]$. Explicitly, $\mathbb{E}[A \circ B] = (1 \pm 80\varepsilon)\,\mathbb{E}[A]\,\mathbb{E}[B] \pm 2^{-k}$.

2.4 Cylinder intersection closure

We now turn back to prove Theorem 2.4. It will be more convenient to prove an equivalent version of it, which relates to the *cylinder intersection closure* of sets.

Definition 2.9 (Cylinder intersection closure)**.** *Let $T \subset [N]^3$. Its cylinder intersection closure, denoted $\overline{CI}(T)$, is the smallest cylinder intersection containing T. Explicitly, if we denote the marginals of T by*

$$T_{XY} = \{(x,y) : (x,y,z) \in T\}, \quad T_{XZ} = \{(x,z) : (x,y,z) \in T\}, \quad T_{YZ} = \{(y,z) : (x,y,z) \in T\},$$

then its cylinder intersection closure is

$$\overline{CI}(T) = CI(T_{XY}, T_{XZ}, T_{YZ}).$$

In other words, $\overline{CI}(T)$ is the set of all points (x,y,z) such that $(x,y,z'), (x,y',z), (x',y,z) \in T$ for some $x', y', z' \in [N]$.

The following theorem is equivalent to Theorem 2.4, and more convenient for us to prove (see Claim 5.1 for a formal proof of equivalence).

Theorem 2.10. *Let $D \subset [N]^3$ be γ-pseudorandom with respect to large cubes. Let $T \subset D$ of size $|T| \geq 2^{-d}|D|$, and assume $\gamma \leq 2^{-O(d^3)}$. Then*

$$|\overline{CI}(T)| \geq 2^{-O(d^3)} N^3.$$

We now focus on proving Theorem 2.10. We first fix some notations and conventions. We will consider sets within cubes, $T \subset X \times Y \times Z$. Given such a set, and a function $f : X \times Y \to [0,1]$, we say that the function f is supported on the XY marginal of T if $\text{supp}(f) \subset T_{XY}$. We similarly define when a function is supported on the XZ, YZ marginals of f.

We define a notion of a "well behaved" set T inside a cube – a property which depends only on the marginals of T. We will show that for such sets, their cylinder intersection closure must be large.

Definition 2.11 (Well-behaved sets)**.** *Let $T \subset X \times Y \times Z$, and let $d \geq 1, r \geq 1, \varepsilon \in (0,1)$. We say that T is (d,r,ε)-well behaved if there exist bounded functions $f : X \times Z \to [0,1]$, $g : Y \times Z \to [0,1]$, $h : X \times Y \to [0,1]$, supported on the XZ, YZ, and XY-marginals of T, respectively, such that the following conditions hold:*

1. $\mathbb{E}[f], \mathbb{E}[g], \mathbb{E}[h] \geq 2^{-d}$.
2. *f, g are (r,ε)-spread.*
3. *f, g are ε-left-lower bounded.*

The following lemma shows that the cylinder intersection closure of well-behaved sets is large. Its proof is a direct application of Theorem 2.8.

Lemma 2.12 (Well behaved sets have large cylinder intersection closure)**.** *Let $T \subset X \times Y \times Z$. Assume that T is (d,r,ε)-well behaved for $d \geq 1$, $r = \Omega(d^2), \varepsilon = O(1)$. Then*

$$|\overline{CI}(T)| \geq 2^{-O(d)} |X||Y||Z|.$$

Typically, given a set $T \subset [N]^3$, it will not be well-behaved. However, if there exists a large cube C such that $T \cap C$ (considered as a subset of the cube C) is well-behaved, then we may apply Lemma 2.12 to $T \cap C$ and obtain that its cylinder intersection closure is large, and hence the same holds for T. The next lemma guarantees that such a cube exists.

Lemma 2.13 (Finding well-behaved sets). *Let $D \subset [N]^3$ be γ-pseudorandom with respect to large cubes. Let $d \geq 1, r \geq 1, \varepsilon \in (0,1)$, and assume $\gamma \leq 2^{-\Omega(dr/\varepsilon)}$. Let $T \subset D$ of size $|T| \geq 2^{-d}|D|$. Then there is a cube $C \subset [N]^3$ of size $|C| \geq 2^{-O(dr/\varepsilon)}N^3$ such that $T \cap C$ (considered as a subset of the cube C) is $(d+2, r, \varepsilon)$-well behaved.*

Combining Lemma 2.12 and Lemma 2.13 proves Theorem 2.10, which to recall is equivalent to Theorem 2.4.

Proof of Theorem 2.10. Let $T \subset D$ of size $|T| \geq 2^{-d}|D|$. Apply Lemma 2.13 with $r = O(d^2), \varepsilon = O(1)$ which we may since we assume $\gamma \leq 2^{-\Omega(d^3)}$. This implies the existence of a cube $C \subset [N]^3$ of size $|C| \geq 2^{-O(d^3)}N^3$ such that $T \cap C$ (considered as a subset of the cube C) is $(d+2, r, \varepsilon)$-well behaved. Lemma 2.12 then gives that

$$|\overline{CI}(T)| \geq |\overline{CI}(T \cap C)| \geq 2^{-O(d)}|C| \geq 2^{-O(d^3)}N^3.$$

□

2.5 Construction of sparse pseudorandom sets

The last piece of the puzzle is constructing sets $D \subset [N]^3$ which are pseudorandom with respect to large cubes, sparse, and at the same time easy for randomized NOF protocols. Recall that we identify a set D with its indicator function $D : [N]^3 \to \{0,1\}$, which will be the function for which we consider NOF protocols.

In order to define such sets D, we first define the notion of *expander-colorings*, which are colorings of the edges of the complete bi-partite graph such that each color class is a good expander (it is a variant of two-source extractors). We will then use them to construct the desired set D.

Definition 2.14 (Expander-coloring). *Let $Col : [N] \times [N] \to [M]$. We say that Col is an (N, M, η)-expander coloring if for any sets $X, Y \subset [N]$ of size $|X|, |Y| \geq \eta N$ and any color $m \in [M]$,*

$$|\{(x,y) \in X \times Y : Col(x,y) = m\}| = \frac{|X||Y|}{M}(1 \pm \eta).$$

We use expander-colorings to define the desired set $D \subset [N]^3$. Given $\mathrm{Col} : [N] \times [N] \to [M]$, we define its corresponding set $D(\mathrm{Col}) \subset [N]^3$ as the set of triples, such that the color of each of the pairs is the same:

$$D(\mathrm{Col}) = \{(x,y,z) \in [N]^3 : \mathrm{Col}(x,y) = \mathrm{Col}(x,z) = \mathrm{Col}(y,z)\}.$$

We note that for any expander-coloring Col, the associated function $D(\mathrm{Col})$ is easy for randomized NOF protocols, since membership in D can be decided by checking the pairwise equalities between $\mathrm{Col}(x,y)$, $\mathrm{Col}(x,z)$, and $\mathrm{Col}(y,z)$.

Claim 2.15. *For any $Col : [N] \times [N] \to [M]$, there is a randomized NOF protocol (using public randomness) which sends $O(1)$ bits and computes $D(Col)$.*

We show that if Col is a good expander-coloring, then its associated set $D(\mathrm{Col})$ is pseudorandom with respect to large cubes.

Lemma 2.16 (Pseudorandom sets from expander-colorings)**.** *Let $Col : [N] \times [N] \to [M]$ be an (N, M, η)-expander coloring. Let $\gamma \in (0, 1)$, and assume $\eta = O(\gamma^3 M^{-5})$. Then $D(Col)$ is γ-pseudorandom with respect to large cubes.*

To conclude, we need an explicit construction of an expander-coloring with good parameters. We show that the inner-product function is a good expander-coloring.

Lemma 2.17 (Expander-coloring based on inner-product)**.** *Let q be a prime power and $k \geq 3$. Consider the inner-product function $IP : \mathbb{F}_q^k \times \mathbb{F}_q^k \to \mathbb{F}_q$ given by*

$$IP(x, y) = \langle x, y \rangle = \sum_{i=1}^{k} x_i y_i.$$

Then IP is a (q^k, q, η)-expander coloring for $\eta = q^{-(k-2)/4}$.

We can now combine together all the machinery we developed, and prove our main theorem, Theorem 1.1.

Proof of Theorem 1.1. Let q be a prime power and $k = 34$ (any larger constant would also work). Lemma 2.17 gives that the inner-product function IP on $\mathbb{F}_q^k$ is a (N, M, η)-expander coloring for $N = q^k, M = q, \eta = q^{-8}$. Lemma 2.16 gives that $D = D(\mathrm{IP})$ is γ-pseudorandom with respect to large cubes for $\gamma = O(q^{-1})$. In particular, this implies that $|D| = \Theta(N^3 q^{-2})$, so D is sparse, concretely $|D| = \Theta(N^{3-c})$ for $c = 2/3k$. We may thus apply Theorem 2.2 and obtain that the non-deterministic NOF complexity of D is $\Omega((\log N)^{1/3})$. Finally, Claim 2.15 shows that the randomized NOF complexity of D is $O(1)$. □

Paper organization. We start in Section 3 with some preliminary definitions. The proofs of the lemmas from Section 2.3, Section 2.4 and Section 2.5 are given in Section 4, Section 5 and Section 6, respectively. We discuss some open problems in Section 7.

Acknowledgements. We thank Sreejato Bhattacharya who found a small bug in a previous version of this paper.

3 Preliminaries

Notations. Given positive numbers x, y, we shorthand $x = y \pm \varepsilon$ for $y - \varepsilon \leq x \leq y + \varepsilon$. We similarly define $x = y(1 \pm \varepsilon)$. Given $x \in \mathbb{R}$ we denote $x_+ = \max(x, 0)$ and $x_- = \max(-x, 0)$ so that $x = x_+ - x_-$. Given a function $f : \Omega \to \mathbb{R}$ for some domain Ω, its support is $\mathrm{supp}(f) = \{x \in \Omega : f(x) \neq 0\}$.

Norms and normalizations. Given a real-valued random variable X, its k-th norm is $\|X\|_k := \mathbb{E}[|X|^k]^{1/k}$. Similarly, for a real-valued function $f : \Omega \to \mathbb{R}$ defined on some ambient finite set Ω, its k-th norm is $\|f\|_k := \left(\mathbb{E}_{x \in \Omega} |f(x)|^k\right)^{1/k}$. Given two functions $f, g : \Omega \to \mathbb{R}$, their inner product is $\langle f, g \rangle := \mathbb{E}_{x \in \Omega}[f(x)g(x)]$.

We need the following claim from [KM23].

Claim 3.1 ([KM23]). *Let $\varepsilon \in [0, \frac{1}{4}]$. Suppose X is a real-valued random variable with*

- $\|X\|_k \geq 2\varepsilon$ *for some even* $k \in \mathbb{N}$,
- $\mathbb{E}\, X^j \geq 0$ *for all* $j \in \mathbb{N}$.

Then $\|1 + X\|_p \geq 1 + \varepsilon$ *for all integers* $p \geq k/\varepsilon$.

Communication complexity. We consider NOF 3-party protocols in three communication models: deterministic, randomized (with public ranodmness), and non-deterministic. They are the analog models for the corresponding two-party models. Formally, the deterministic NOF communication complexity of f is the least number of bits needed to compute the function by a deterministic NOF protocol; the randomized NOF communication complexity of f is the least number of bits needed to compute the function by a randomized NOF protocol, with error at most $1/3$, and with access to public randomness; and the non-deterministic NOF communication complexity of f is the least number of bits needed to compute f by a non-deterministic NOF protocol. For more details see any textbook on communication complexity, for example [KN96] or [RY20].

4 Uniformity from spreadness

In this section we prove our main technical tool, Theorem 2.8, which says roughly that for two spread matrices A and B, the product $A \circ B$ is uniform. There are two main steps to the argument, and there is a strong analogy between these steps with the arguments involved in the "sifting" and "spectral positivity" sections of [KM23]. However, the proof we give will be self-contained.[3]

4.1 Grid norms

We first discuss the following quantity (the "grid norm" of a matrix) which is central to the argument. This quantity is particularly useful for succinctly capturing certain kinds of double-counting arguments involving dense bipartite graphs. For example, the proof of Lemma 7.4 in [Gow01] involves (implicitly) the quantity $U_{2,5}(A)$, where A is the adjacency matrix encoding the set system involved. In addition, it is known in various contexts that this quantity can be reasonably interpreted as some kind of measure of pseudorandomness (A is "pseudorandom" when $\|A\|_{U(\ell,k)}$ is not much larger than $\|A\|_{U(1,1)}$) – see e.g. [Gow06, Theorem 3.1] which discusses the approximate equivalence of a bound on $\|A\|_{U(2,2)}$ with some other measures of pseudorandomness.

Definition 4.1 (Grid norms). *For a function $f : X \times Z \to \mathbb{R}$, and $\ell, k \in \mathbb{N}$, let*

$$\begin{aligned} U_{\ell,k}(f) &:= \mathop{\mathbb{E}}_{x_1,x_2,\ldots,x_\ell \in X} \left(\mathop{\mathbb{E}}_{z \in Z} f(x_1, z) f(x_2, z) \cdots f(x_\ell, z) \right)^k \\ &= \mathop{\mathbb{E}}_{z_1,z_2,\ldots,z_k \in Z} \left(\mathop{\mathbb{E}}_{x \in X} f(x, z_1) f(x, z_2) \cdots f(x, z_k) \right)^\ell \\ &= \mathop{\mathbb{E}}_{\substack{x \in X^\ell \\ z \in Z^k}} \prod_{i=1}^{\ell} \prod_{j=1}^{k} f(x_i, z_j). \end{aligned}$$

We also write

$$\|f\|_{U(\ell,k)} := |U_{\ell,k}(f)|^{1/\ell k}.$$

[3] That is, self-contained except for the use of a basic fact about random variables with non-negative odd moments, Claim 3.1.

The name "grid norm" is a bit of a misnomer, as $\|\cdot\|_{U(\ell,k)}$ is not a norm in general; still, the name captures the essence in which we use them. It is known that $\|\cdot\|_{U(\ell,k)}$ is a semi-norm (that is, it satisfies a triangle inequality) whenever ℓ, k are both even [Hat10, Theorems 2.8, 2.9] – a fact which we don't use in this work.

We record some basic properties of $\|\cdot\|_{U(\ell,k)}$.

Claim 4.2 (Monotonicity)**.** *Let $\ell, k, \ell', k' \in \mathbb{N}$, where $\ell \le \ell'$, $k \le k'$. Let $A : X \times Z \to \mathbb{R}_{\ge 0}$. Then,*

$$\|A\|_{U(\ell,k)} \le \|A\|_{U(\ell',k)} \quad \text{and} \quad \|A\|_{U(\ell,k)} \le \|A\|_{U(\ell,k')}.$$

Proof. By symmetry it suffices to show that $\|A\|_{U(\ell,k)} \le \|A\|_{U(\ell,k')}$. So, fix $\ell \in \mathbb{N}$. For uniformly random $x \in [N]^\ell$, consider the resulting random variable

$$D = D(x) := \left(\mathop{\mathbb{E}}_{z \in Z} A(x_1, z) A(x_2, z) \cdots A(x_\ell, z) \right)^k.$$

Note that D is non-negative since we assume A is non-negative. We have

$$\mathbb{E}\, D \le (\mathbb{E}\, D^r)^{1/r}$$

for any $r \ge 1$, which for $r = k'/k$ shows that

$$\|A\|_{U(\ell,k)} = (\mathbb{E}\, D)^{1/\ell k} \le (\mathbb{E}\, D^{k'/k})^{1/\ell k'} = \|A\|_{U(\ell,k')}. \qquad \square$$

Lemma 4.3 (Decoupling inequality for $U_{2,k}$)**.** *Let $f : X \times Z \to \mathbb{R}$ and $g : Y \times Z \to \mathbb{R}$. For even $k \in \mathbb{N}$ we have*

$$\mathop{\mathbb{E}}_{x,y} \left(\mathop{\mathbb{E}}_{z} f(x,z) g(y,z) \right)^k \le U_{2,k}(f)^{1/2} \cdot U_{2,k}(g)^{1/2}.$$

Proof. Let $x \in X, y \in Y, z_1, \ldots, z_k \in Z$. Then

$$\begin{aligned}
\mathop{\mathbb{E}}_{x,y} \left(\mathop{\mathbb{E}}_{z} f(x,z) g(y,z) \right)^k &= \mathop{\mathbb{E}}_{x,y} \mathop{\mathbb{E}}_{z_1,\ldots,z_k} \left[\prod_{i=1}^k f(x, z_i) \cdot \prod_{i=1}^k g(y, z_i) \right] \\
&= \mathop{\mathbb{E}}_{z_1,\ldots,z_k} \mathop{\mathbb{E}}_{x} \left[\prod_{i=1}^k f(x, z_i) \right] \cdot \mathop{\mathbb{E}}_{y} \left[\prod_{i=1}^k g(y, z_i) \right] \\
&\le \sqrt{\mathop{\mathbb{E}}_{z_1,\ldots,z_k} \left(\mathop{\mathbb{E}}_{x} \prod_{i=1}^k f(x, z_i) \right)^2} \sqrt{\mathop{\mathbb{E}}_{z_1,\ldots,z_k} \left(\mathop{\mathbb{E}}_{y} \prod_{i=1}^k g(y, z_i) \right)^2} \\
&= \sqrt{U_{2,k}(f)} \sqrt{U_{2,k}(g)}.
\end{aligned} \qquad \square$$

Claim 4.4. *Let $f : X \times Z \to \mathbb{R}$, and consider*

$$M(x, x') := \mathop{\mathbb{E}}_{z \in Z} f(x,z) f(x',z).$$

For any integer $j \in \mathbb{N}$ we have

$$\mathop{\mathbb{E}}_{x,x' \in X} M(x,x')^j \ge 0.$$

Proof. We note that

$$\mathbb{E}\, M(x,x')^j = U_{2,j}(f),$$

and it is clear from the definition of $U_{2,j}(f)$ that the quantity is non-negative for all real-valued functions f:

$$U_{2,j}(f) = \mathop{\mathbb{E}}_{z_1,z_2,\ldots,z_j \in Z} \left(\mathop{\mathbb{E}}_{x\in X} f(x,z_1)f(x,z_2)\cdots f(x,z_k) \right)^2 \geq 0. \qquad \square$$

4.2 Spread matrices have controlled grid norms

Let

$$\mathcal{R} = \left\{ \mathbb{1}_{X'}(x)\mathbb{1}_{Y'}(y) \ : \ X' \times Y' \subseteq X \times Y \right\}$$

denote the set of rectangle indicator functions, and let

$$\mathrm{conv}(\mathcal{R}) = \left\{ \sum_i c_i \mathbb{1}_{R_i} \ : \ \mathbb{1}_{R_i} \in \mathcal{R}, c_i \geq 0, \sum_i c_i \leq 1 \right\}$$

denote its convex hull. We also consider the slightly richer class of "soft rectangles",

$$\widetilde{\mathcal{R}} = \left\{ f(x)g(y) \ : \ f : X \to [0,1], g : Y \to [0,1] \right\}.$$

Next, we note that any soft rectangle can be expressed as a convex combination of rectangles. In particular, this implies that $\mathrm{conv}(\widetilde{\mathcal{R}}) = \mathrm{conv}(\mathcal{R})$.

Claim 4.5. *$\widetilde{\mathcal{R}} \subseteq \mathrm{conv}(\mathcal{R})$.*

Proof. Let $f(x)g(y)$ be a soft rectangle. We can write[4]

$$f(x)g(y) = \int_{t=0}^{1} \int_{s=0}^{1} \mathbb{1}(f(x) \geq s)\, \mathbb{1}(g(y) \geq t)\, ds\, dt. \qquad \square$$

The next lemma says that any non-negative function $D : X \times Y \to \mathbb{R}_{\geq 0}$ that has a non-trivial correlation with an element of $\mathrm{conv}(\mathcal{R})$ also does so with a rectangle of comparable density.

Claim 4.6. *Let $D : X \times Y \to \mathbb{R}_{\geq 0}$ and $F \in \mathrm{conv}(\mathcal{R})$; suppose that $\|D\|_\infty \leq \Delta$ and $\|F\|_1 \geq \delta$. If*

$$\left\langle \frac{F}{\|F\|_1}, D \right\rangle \geq 1 + \varepsilon,$$

then there is some rectangle R with

$$\mathop{\mathbb{E}}_{(x,y)\in R} D(x,y) = \left\langle \frac{\mathbb{1}_R}{\|\mathbb{1}_R\|_1}, D \right\rangle \geq 1 + \frac{\varepsilon}{2}$$

and

$$\frac{|R|}{|X||Y|} = \|1_R\|_1 \geq \frac{\varepsilon\delta}{2\Delta}.$$

[4] If desired, one may obtain a representation as a finite combination of rectangles by noting that there are only finitely many different superlevel sets $\{x \ : \ f(x) \geq s\}$ and $\{y \ : \ g(y) \geq t\}$.

Proof. Write $F = \sum_i c_i \mathbb{1}_{R_i}$. We begin by pruning rectangles which are too small: define $F' = \sum_i c_i' \mathbb{1}_{R_i}$ via

$$c_i' = \begin{cases} c_i & \text{if } \|\mathbb{1}_{R_i}\|_1 \geq \tau, \\ 0 & \text{if } \|\mathbb{1}_{R_i}\|_1 < \tau \end{cases}$$

for some threshold value τ. We note that

$$\frac{\langle F', D\rangle}{\|F\|_1} = \frac{\langle F, D\rangle}{\|F\|_1} - \frac{\langle F - F', D\rangle}{\|F\|_1} \geq 1 + \varepsilon - \frac{\|F - F'\|_1 \|D\|_\infty}{\|F\|_1} \geq 1 + \varepsilon - \frac{\tau\Delta}{\delta}.$$

We set $\tau = \frac{\varepsilon\delta}{2\Delta}$, giving

$$\frac{\langle F', D\rangle}{\|F'\|_1} \geq \frac{\langle F', D\rangle}{\|F\|_1} \geq 1 + \frac{\varepsilon}{2}.$$

In particular, we must have $\langle F', D\rangle > 0$, which guarantees that F' is not identically zero. We have

$$\frac{\langle F', D\rangle}{\|F'\|_1} = \frac{\sum_i c_i' \langle \mathbb{1}_{R_i}, D\rangle}{\sum_i c_i' \|\mathbb{1}_{R_i}\|_1}.$$

By averaging, there is some choice of $R = R_i$ with

$$\left\langle \frac{\mathbb{1}_R}{\|\mathbb{1}_R\|_1}, D \right\rangle \geq 1 + \frac{\varepsilon}{2}$$

and

$$\|\mathbb{1}_R\|_1 \geq \tau. \qquad \square$$

Lemma 4.7. *(Sifting a rectangle) Let $A : X \times Y \to [0,1]$; suppose that $\|A\|_1 \geq \delta$. Let $\ell, k \in \mathbb{N}$. If*

$$\|A\|_{U(\ell,k)} \geq (1+\varepsilon)\|A\|_1,$$

then there is some rectangle R with

$$\mathop{\mathbb{E}}_{(x,y)\in R} A(x,y) \geq \left(1 + \frac{\varepsilon}{2}\right) \|A\|_1$$

and

$$\frac{|R|}{|X||Y|} = \|\mathbb{1}_R\|_1 \geq \tfrac{1}{2} \cdot \varepsilon \cdot \delta^{\ell k+1}.$$

In particular, if A is (r, ε)-spread for some $r \geq (\ell k + 1)\log(1/\delta) + \log(1/\varepsilon)$, then

$$\|A\|_{U(\ell,k)} \leq (1 + 2\varepsilon)\|A\|_1.$$

Proof. By assumption we have

$$\|A\|_{U(\ell,k)}^{\ell k} = \mathop{\mathbb{E}}_{\substack{x_1,\ldots,x_\ell \in X \\ y_1,\ldots,y_k \in Y}} \left[\prod_{i=1}^{\ell} \prod_{j=1}^{k} A(x_i, y_j) \right] \geq (1+\varepsilon)^{\ell k} \|A\|_1^{\ell k}.$$

Our task is to find a reasonably large rectangle R where A is notably denser than average. Before proceeding with the actual argument, let us offer a not-entirely-accurate picture of how this will be done. For illustration, suppose A is an adjacency matrix of a bipartite graph with parts X and Y. We then look for the desired rectangle among those of the following specific form. For any choice of x_i's and y_j's we can consider the rectangle $R = \Gamma(y_1, y_2, \ldots, y_k) \times \Gamma(x_1, x_2, \ldots, x_\ell) \subseteq X \times Y$, where (e.g.) $\Gamma(y_1, y_2, \ldots, y_k) \subseteq X$ denotes the set of common neighbors of the vertices $y_1, \ldots, y_k$ within our bipartite graph. In the actual argument, we will need to consider also some additional, related choices for R which are not so nicely describable. We then use our assumption on $\|A\|_{U(\ell,k)}$ to argue that one of these choices must succeed. We now proceed with the argument (considering now general $A : X \times Y \to [0,1]$).

Let us fix some arbitrary ordering on tuples $(i,j) \in [\ell] \times [k]$ (say, the lexicographic ordering), and consider the prefix-products

$$\phi_{\le(i,j)}(A) := \prod_{(i',j')\le(i,j)} A(x_{i'}, y_{j'}).$$

Thus, we have $\mathbb{E}[\phi_{\le(1,1)}(A)] = \|A\|_1$ and $\mathbb{E}[\phi_{\le(\ell,k)}(A)] = \|A\|_{U(\ell,k)}^{\ell k}$. Similarly, let us write

$$\phi_{<(i,j)}(A) := \prod_{(i',j')<(i,j)} A(x_{i'}, y_{j'}),$$

with the convention $\phi_{<(1,1)}(A) := 1$. Now consider the telescoping product

$$\prod_{(i,j)\in[\ell]\times[k]} \frac{\mathbb{E}\,\phi_{\le(i,j)}(A)}{\mathbb{E}\,\phi_{<(i,j)}(A)} = \|A\|_{U(\ell,k)}^{\ell k}$$

This quantity is at least $(1+\varepsilon)^{\ell k}\|A\|_1^{\ell k}$, and so we infer that for some choice of (i^*, j^*) we have

$$\frac{\mathbb{E}\,\phi_{\le(i^*,j^*)}(A)}{\mathbb{E}\,\phi_{<(i^*,j^*)}(A)} \ge (1+\varepsilon)\|A\|_1.$$

At this point we would like to think of $\phi_{<(i^*,j^*)}(A)$ primarily as a function of x_{i^*} and y_{j^*}. Let us define

$$F(x_{i^*}, y_{j^*}) = \mathop{\mathbb{E}}_{\substack{x_1,\ldots,x_{i^*-1},x_{i^*+1},\ldots,x_\ell\in X\\ y_1,\ldots,y_{j^*-1},y_{j^*+1},\ldots,y_k\in Y}} \prod_{(i,j)<(i^*,j^*)} A(x_i, y_j)$$

so that we may reinterpret

$$\mathbb{E}\,\phi_{\le(i^*,j^*)}(A) = \mathbb{E}\,\phi_{<(i^*,j^*)}(A)\cdot A(x_{i^*}, y_{j^*}) = \mathop{\mathbb{E}}_{\substack{x_{i^*}\in X\\ y_{j^*}\in Y}} F(x_{i^*}, y_{j^*})\cdot A(x_{i^*}, y_{j^*}) = \langle F, A\rangle$$

and

$$\mathbb{E}\,\phi_{<(i^*,j^*)}(A) = \mathbb{E}[F] = \|F\|_1.$$

Thus, we have

$$\left\langle \frac{F}{\|F\|_1}, \frac{A}{\|A\|_1} \right\rangle \ge 1+\varepsilon.$$

Finally, we argue that F is a convex combination of soft rectangles so that we may finish the proof by applying Claim 4.6 (to F and $D := A/\|A\|_1$). Indeed, as a function of x_{i^*}, y_{j^*}, and for any fixing of the other variables x_i, y_j, the quantity

$$\prod_{(i,j)<(i^*,j^*)} A(x_i,y_j) = \left(\prod_{\substack{(i,j)<(i^*,j^*) \\ i\neq i^* \\ j\neq j^*}} A(x_i,y_j) \right) \left(\prod_{\substack{(i,j)<(i^*,j^*) \\ i=i^*}} A(x_{i^*},y_j) \right) \left(\prod_{\substack{(i,j)<(i^*,j^*) \\ j=j^*}} A(x_i,y_{j^*}) \right)$$

is a soft rectangle: each of the factors $A(x_i,y_j)$ may depend on x_{i^*} or y_{j^*} but not both, and the product of 1-bounded functions is again a 1-bounded function.

It remains only to discuss what sort of bounds we have on $\|F\|_1$ and $\|D\|_\infty$. Since A is 1-bounded, we have $\|D\|_\infty \leq \frac{1}{\|A\|_1} \leq \frac{1}{\delta}$. It follows also from the 1-boundedness of A that

$$\|F\|_1 = \mathbb{E}\,\phi_{<(i^*,j^*)} \geq \mathbb{E}\,\phi_{\leq(\ell,j)} = \|A\|_{U(\ell,k)}^{\ell k} \geq \delta^{\ell k}.$$

Thus, Claim 4.6 provides a rectangle $R = X' \times Y'$ of the desired size, $\|\mathbb{1}_R\|_1 \geq \varepsilon\delta^{\ell k+1}/2$. □

4.3 Products of $U_{2,k}$-regular matrices

Lemma 4.8 ($U_{2,k}$-regularity of A and B implies uniformity of $A \circ B$)**.** *Fix an even integer $k \in \mathbb{N}$, $\varepsilon \in (0, 1/20)$, and set $p = \lceil k/\varepsilon \rceil$. Let $A : X \times Z \to \mathbb{R}_{\geq 0}$, $B : Y \times Z \to \mathbb{R}_{\geq 0}$ be two (nonzero) matrices, and suppose that*

- $\|A\|_{U(2,p)} \leq (1+\varepsilon)\|A\|_1$,
- $\|B\|_{U(2,p)} \leq (1+\varepsilon)\|B\|_1$,
- *A, B are ε-left lower bounded.*

Then, the function $D(x,y) = \frac{(A\circ B)(x,y)}{\mathbb{E}[A]\,\mathbb{E}[B]}$ is close to uniform on $X \times Y$:

$$\|D - 1\|_k \leq 20\varepsilon.$$

Proof. Note that the statement is scale-invariant with respect to A and B. Without loss of generality suppose that $\mathbb{E}[A] = \mathbb{E}[B] = 1$. For brevity, let $A_x, B_y : Z \to \mathbb{R}$ be the functions $A_x(z) = A(x,z), B_y(z) = B(y,z)$. For any two functions $\alpha, \beta : Z \to \mathbb{R}$, we write $\langle \alpha, \beta \rangle := \mathbb{E}_z[\alpha(z)\beta(z)]$. With this notation we can express

$$D(x,y) = \langle A_x, B_y \rangle\,.$$

For what follows it will be convenient to introduce the notation $a(x) := \mathbb{E}_z\, A(x,z)$ and $b(y) := \mathbb{E}_z\, B(y,z)$.

We proceed to argue that $\|\langle A_x, B_y\rangle - 1\|_k \leq O(\varepsilon)$. We have

$$\langle A_x, B_y \rangle - 1 = \langle A_x - 1, B_y - 1 \rangle + \langle A_x - 1, 1 \rangle + \langle B_y - 1, 1 \rangle\,.$$

We first consider the latter terms. Note that

$$\|\langle A_x - 1, 1 \rangle\|_k = \|a-1\|_k \leq \|(a-1)_-\|_k + \|(a-1)_+\|_k.$$

We handle the positive and negative deviations from 1 separately. First, since we assume A is ε-left lower bounded we have that $a(x) \geq 1-\varepsilon$ pointwise, which gives

$$\|(a-1)_-\|_k \leq \|(a-1)_-\|_\infty \leq \varepsilon.$$

Second, we note that for uniformly random x, the resulting random variable $(a(x)-1)_+$ certainly has non-negative odd moments since it is non-negative, and so with Claim 3.1 we can obtain a bound on $\|(a-1)_+\|_k$ from a bound on $\|1+(a-1)_+\|_p$. Specifically, we have $1+(a-1)_+ = a+(a-1)_-$ and hence

$$\|1+(a-1)_+\|_p \leq \|a\|_p + \|(a-1)_-\|_p \leq \|a\|_p + \varepsilon$$

and

$$\|a\|_p = \|A\|_{U(1,p)} \leq \|A\|_{U(2,p)} \leq 1+\varepsilon.$$

We conclude that $\|1+(a-1)_+\|_p \leq 1+2\varepsilon$, and hence by Claim 3.1

$$\|(a-1)_+\|_k \leq 4\varepsilon.$$

Overall, we obtain the bound $\|a-1\|_k \leq 5\varepsilon$. Similarly, also $\|b-1\|_k \leq 5\varepsilon$. Therefore, by the triangle inequality for $\|\cdot\|_k$, we have

$$\|\langle A_x, B_y\rangle - 1\|_k \leq \|\langle A_x-1, B_y-1\rangle\|_k + 10\varepsilon.$$

We now apply the decoupling inequality for $U_{2,k}$ (Lemma 4.3) to study the main term.

$$\|\langle A_x-1, B_y-1\rangle\|_k^k = \mathop{\mathbb{E}}_{x,y}\left[\mathop{\mathbb{E}}_{z}\Big((A(x,z)-1)(B(y,z)-1)\Big)^k\right] \leq U_{2,k}(A-1)^{1/2}\, U_{2,k}(B-1)^{1/2},$$

or equivalently,

$$\|\langle A_x-1, B_y-1\rangle\|_k \leq \|A-1\|_{U(2,k)}\|B-1\|_{U(2,k)}.$$

Without loss of generality, suppose $U_{2,k}(A-1) \geq U_{2,k}(B-1)$ so that $\|A-1\|^2_{U(2,k)} \geq \|\langle A_x-1, B_y-1\rangle\|_k$. Seeking contradiction, we observe that if $\|\langle A_x, B_y\rangle - 1\|_k > 20\varepsilon$, then $\|\langle A_x-1, B_y-1\rangle\|_k > 10\varepsilon$, and so

$$\|A-1\|^2_{U(2,k)} > 10\varepsilon.$$

Let $M(x,x') = \langle A_x-1, A_{x'}-1\rangle$, and note that

$$U_{2,k}(A-1) = \mathop{\mathbb{E}}_{x,x'}\left(\langle A_x-1, A_{x'}-1\rangle\right)^k = \|M\|_k^k.$$

Consider the random variable $M = M(x,x')$ arising from a uniform random choice of $x, x' \in X$. As observed in Claim 4.4, M has non-negative odd moments. Therefore, by Claim 3.1, $\|1+M\|_p > 1+5\varepsilon$. Further,

$$\begin{aligned}\langle A_x, A_{x'}\rangle &= 1 + \langle A_x-1, 1\rangle + \langle 1, A_{x'}-1\rangle + M(x,x')\\ &= 1 + M(x,x') + (a(x)-1) + (a(x')-1)\\ &\geq 1 + M(x,x') - (a(x)-1)_- - (a(x')-1)_-,\end{aligned}$$

and so

$$\begin{aligned}
\| \langle A_x, A_{x'} \rangle \|_p &\geq \|1 + M\|_p - 2\|(a-1)_-\|_p \\
&\geq \|1 + M\|_p - 2\varepsilon \\
&> 1 + 3\varepsilon,
\end{aligned}$$

since we already noted that $\|(a-1)_-\|_\infty \leq \varepsilon$. Thus, we have $\| \langle A_x, A_{x'} \rangle \|_p > 1 + 3\varepsilon$. On the other hand, our regularity assumption on A says that

$$\| \langle A_x, A_{x'} \rangle \|_p = \|A\|_{U(2,p)}^2 \leq (1+\varepsilon)^2 < 1 + 3\varepsilon,$$

giving a contradiction. Thus, we must in fact have $\| \langle A_x, B_y \rangle - 1\|_k \leq 20\varepsilon$. □

Proof of Theorem 2.8. The proof is a straightforward combination of Lemma 4.7 with Lemma 4.8, followed by an application of Hölder's inequality. Let $p = \lceil k/\varepsilon \rceil$. We would like to apply Lemma 4.7 to control $U_{2,p}(A)$ and $U_{2,p}(B)$ with our spreadness assumption, which is indeed possible for $r = cdk/\varepsilon$ where $c > 0$ is a sufficiently large constant. We conclude that

$$\|A\|_{U(2,p)} \leq (1+2\varepsilon)\|A\|_1, \qquad \|B\|_{U(2,p)} \leq (1+2\varepsilon)\|B\|_1.$$

Consider

$$D(x,y) := \frac{(A \circ B)(x,y)}{\mathbb{E}[A]\,\mathbb{E}[B]}.$$

From Lemma 4.8 we have

$$\|D - 1\|_k \leq 40\varepsilon.$$

For any subset $S \subseteq X \times Y$ of size $|S| \geq 2^{-k}|X \times Y|$ we have

$$\begin{aligned}
\mathop{\mathbb{E}}_{(x,y)\in S} |D(x,y) - 1| &\leq \left(\mathop{\mathbb{E}}_{(x,y)\in S} |D(x,y) - 1|^k \right)^{1/k} \\
&\leq \left(2^k \mathop{\mathbb{E}}_{(x,y)\in X\times Y} |D(x,y) - 1|^k \right)^{1/k} \\
&= 2\|D - 1\|_k \\
&\leq 80\varepsilon.
\end{aligned}$$

If we consider the set $T \subseteq X \times Y$ of large deviations

$$T = \{(x,y) \;:\; |D(x,y) - 1| > 80\varepsilon\},$$

it must be the case that $|T| < 2^{-k}|X \times Y|$. Otherwise, we would obtain a large set T with

$$\mathop{\mathbb{E}}_{(x,y)\in T} |D(x,y) - 1| > 80\varepsilon,$$

contradicting the calculation above. □

4.4 Correlations involving spread matrices

Before continuing we record the following (immediate) corollary of Theorem 2.8: if f and g are functions which are dense, spread, and left-lower bounded, and h is any function which is dense, then the quantity $\mathbb{E}_{x,y,z}\, f(x,z)g(y,z)h(x,y)$ behaves roughly as if the three terms were independent.

Corollary 4.9. *Let $f : X \times Z \to [0,1]$, $g : Y \times Z \to [0,1]$, $h : X \times Y \to [0,1]$. Let $d \geq 1$ and $\delta \in (0,1)$, and set $r = \Omega((d + \log(1/\delta))d/\delta)$ and $\varepsilon = \delta/160$. Assume that:*

1. *$\mathbb{E}[f], \mathbb{E}[g], \mathbb{E}[h] \geq 2^{-d}$.*
2. *f, g are (r, ε)-spread.*
3. *f, g are ε-lower bounded.*

Then

$$\mathop{\mathbb{E}}_{x\in X, y\in Y, z\in Z} [f(x,z)g(y,z)h(x,y)] = (1 \pm \delta)\, \mathbb{E}[f]\, \mathbb{E}[g]\, \mathbb{E}[h].$$

Proof. Define

$$S = \{(x,y) \in X \times Y : (f \circ g)(x,y) = (1 \pm \delta/2)\, \mathbb{E}[f]\, \mathbb{E}[g]\}\,.$$

Applying Theorem 2.8 (with $A = f, B = g$, $\varepsilon = \delta/160$, and $k = 3d + \log(1/\delta) + 1$) gives that $|S| \geq (1 - 2^{-k})|X||Y|$. For $(x,y) \in S$ we have

$$\mathop{\mathbb{E}}_{z\in Z} [f(x,z)g(y,z)h(x,y)] = (f \circ g)(x,y)h(x,y) = (1 \pm \delta/2)\, \mathbb{E}[f]\, \mathbb{E}[g]h(x,y).$$

For $(x,y) \notin S$ we can naively bound

$$\mathop{\mathbb{E}}_{z\in Z} [f(x,z)g(y,z)h(x,y)] \in [0,1].$$

Averaging over all $(x,y) \in X \times Y$ gives

$$\mathop{\mathbb{E}}_{x\in X, y\in Y, z\in Z} [f(x,z)g(y,z)h(x,y)] = (1 \pm \delta/2)\, \mathbb{E}[f]\, \mathbb{E}[g]\, \mathbb{E}[h] \pm \Pr[(x,y) \in S].$$

The claim follows by the choice of k, since

$$\Pr[(x,y) \in S] \leq 2^{-k} \leq (\delta/2)2^{-3d} \leq (\delta/2)\, \mathbb{E}[f]\, \mathbb{E}[g]\, \mathbb{E}[h]. \qquad \square$$

5 Cylinder intersection closure

We prove in this section the two lemmas in Section 2.4: Lemmas 2.12 and 2.13, as well as formally show that Theorem 2.4 and Theorem 2.10 are equivalent.

Claim 5.1. *Theorem 2.4 and Theorem 2.10 are equivalent.*

Proof. Briefly, Theorem 2.10 is the contra-positive form of Theorem 2.4. Specifically, suppose Theorem 2.10 is true. Suppose we have the conditions of Theorem 2.4. Now, set $T = F \cap D$ and $d = c_1 t^{1/3}$ for a small enough constant c_1. If $|T| \geq 2^{-d}|D|$, then we must have $|\overline{CI}(T)| \geq 2^{-c_1 c_2 d^3} N^3$ for some constant c_2. This violates the density of F for a suitable constant c_1. The reverse direction follows similarly. $\square$

The proof of Lemma 2.12 is a straightforward application of Corollary 4.9.

Proof of Lemma 2.12. Let $M = \overline{CI}(T)$. Let $f : X \times Z \to [0,1]$, $g : Y \times Z \to [0,1]$, $h : X \times Y \to [0,1]$ given by the condition that T is (d, r, ε)-well behaved. As f, g, h are bounded and supported on the XZ, YZ, and XY-marginals of T, respectively, we have the pointwise lower-bound

$$M(x, y, z) \geq f(x, z)g(y, z)h(x, y).$$

Apply Corollary 4.9 to f, g, h to conclude that

$$\mathop{\mathbb{E}}_{(x,y,z)\in X\times Y\times Z} M(x, y, z) \geq \mathop{\mathbb{E}}_{(x,y,z)\in X\times Y\times Z} f(x, z)g(y, z)h(x, y) \geq 2^{-O(d)}.$$

This implies that $|M| \geq 2^{-O(d)}|X||Y||Z|$. □

We now move to prove Lemma 2.13. We start with the following claim, which shows how the pseudorandomness of D allows to approximate certain averages of ratios that come up in the proof.

Claim 5.2. *Assume that $D \subset [N]^3$ is γ-pseudorandom with respect to a cube $C = X \times Y \times Z$, and let $T \subset D$. Define the function $h : X \times Y \to [0,1]$ given by*

$$h(x, y) = \frac{\mathbb{E}_{z\in Z}[T(x, y, z)]}{\mathbb{E}_{z\in Z}[D(x, y, z)]}.$$

Then

$$\mathbb{E}[h] = \frac{|T \cap C|}{|D \cap C|} \pm O(\gamma^{1/3}).$$

Proof. Let $\mu(C) = |D \cap C|/|C|$. Define $v : X \times Y \to \mathbb{R}_{\geq 0}$ by

$$v(x, y) = \frac{\mathbb{E}_{z\in Z}[D(x, y, z)]}{\mu(C)}.$$

Note that $\mathbb{E}[v] = 1$, and the second moment of v is

$$\mathbb{E}[v^2] = \frac{\mathbb{E}_{(x,y,z,z')\in X\times Y\times Z\times Z}[D(x, y, z)D(x, y, z')]}{\mathbb{E}_{(x,y,z)\in X\times Y\times Z}[D(x, y, z)]^2}.$$

The assumption that D is γ-pseudorandom with respect to C implies that $\mathbb{E}[v^2] = \frac{1\pm\gamma}{1\pm\gamma} \leq 1 + 4\gamma$ and hence $\mathrm{Var}[v] = O(\gamma)$. Define

$$S = \left\{(x, y) \in X \times Y : v(x, y) = 1 \pm \gamma^{1/3}\right\}.$$

Chebyshev's inequality gives that $|S| \geq (1 - O(\gamma^{1/3}))|X||Y|$. For $(x, y) \in S$ we have

$$h(x, y) = \frac{\mathbb{E}_{z\in Z}[T(x, y, z)]}{(1 \pm \gamma^{1/3})\mu(C)} = (1 \pm 2\gamma^{1/3})\frac{\mathbb{E}_{z\in Z}[T(x, y, z)]}{\mu(C)} = \frac{\mathbb{E}_{z\in Z}[T(x, y, z)]}{\mu(C)} \pm O(\gamma^{1/3}),$$

where we used the fact that $\mathbb{E}_{z\in Z}[T(x, y, z)] \leq \mathbb{E}_{z\in Z}[D(x, y, z)] = (1 \pm \gamma^{1/3})\mu(C) \leq 2\mu(C)$. For $(x, y) \notin S$ we naively bound $h(x, y) \in [0, 1]$. Thus we can estimate

$$\mathbb{E}[h] = \frac{\mathbb{E}_{(x,y,z)\in C}[T(x, y, z)]}{\mu(C)} \pm O(\gamma^{1/3}) \pm \Pr[(x, y) \notin S] = \frac{|T \cap C|}{|D \cap C|} \pm O(\gamma^{1/3}).$$

□

We prove Lemma 2.13 in two steps. First, we do a density increment to find a cube in which the set T is "mostly" well-behaved (with respect to some specific candidate functions f, g, h which are obtained by considering marginals of the uniform distribution on T). The only deficiency will be that not all points will be left-lower bounded, but instead only most of them. Then we do a pruning phase to remove the bad points. We start with the necessary definitions.

Definition 5.3 (Mostly left lower-bounded). *Let $f : X \times Y \to [0,1]$, and let $\varepsilon \in (0,1), \beta \in (0,1)$. We say that f is β-mostly ε-left lower-bounded if for at least a $(1-\beta)$-fraction of $x \in X$, it holds that*

$$\mathbb{E}_{y \in Y}[f(x,y)] \geq (1-\varepsilon)\,\mathbb{E}[f].$$

Definition 5.4 (Mostly well-behaved sets). *Let $T \subset X \times Y \times Z$, and let $d \geq 1, r \geq 1, \varepsilon \in (0,1), \beta \in (0,1)$. We say that T is $(d, r, \varepsilon, \beta)$-mostly well behaved if there exist bounded functions $f : X \times Z \to [0,1]$, $g : Y \times Z \to [0,1]$, $h : X \times Y \to [0,1]$, supported on the XZ, YZ, and XY-marginals of T, respectively, such that the following conditions hold: Suppose that*

1. $\mathbb{E}[f], \mathbb{E}[g], \mathbb{E}[h] \geq 2^{-d}$.
2. *f, g are (r, ε)-spread.*
3. *f, g are β-mostly ε-left-lower bounded.*

Lemma 5.5 (Finding mostly well-behaved sets). *Let $D \subset [N]^3$ be γ-pseudorandom with respect to large cubes. Let $d \geq 1, r \geq 1, \varepsilon \in (0,1), \beta \in (0,1)$, and assume $\gamma \leq 2^{-\Omega(dr/\varepsilon)}\beta$. Let $T \subset D$ of size $|T| \geq 2^{-d}|D|$. Then there is a cube $C \subset [N]^3$ of size $|C| \geq 2^{-O(dr/\varepsilon)}N^3$ such that $T \cap C$ (considered as a subset of the cube C) is $(d+1, r, \varepsilon, \beta)$-well behaved.*

Proof. Let $\eta = O(\varepsilon/r)$. Given a cube $C \subset [N]^3$ define the function

$$\phi(C) = \frac{|T \cap C|}{|D \cap C|} \cdot |C|^{\eta}.$$

Let $C = X \times Y \times Z$ be the cube which maximizes $\phi(\cdot)$. We will show that C satisfies the required properties. Define the functions $f : X \times Z \to [0,1]$, $g : Y \times Z \to [0,1]$, $h : X \times Y \to [0,1]$ as follows:

$$f(x,z) = \frac{\mathbb{E}_{y \in Y} T(x,y,z)}{\mathbb{E}_{y \in Y} D(x,y,z)}, \quad g(y,z) = \frac{\mathbb{E}_{x \in X} T(x,y,z)}{\mathbb{E}_{x \in X} D(x,y,z)}, \quad h(x,y) = \frac{\mathbb{E}_{z \in Z} T(x,y,z)}{\mathbb{E}_{z \in Z} D(x,y,z)}.$$

Observe that indeed f, g, h are supported on the XZ, YZ, XY faces of T, respectively; and that since $T \subset D$, the functions f, g, h take values in $[0,1]$. We will prove that f, g, h are all 2^{-d} dense, (r, ε)-spread and β-mostly ε-left lower-bounded. For concreteness, we prove these properties for h, but they hold for f, g by an analogous argument.

Large cube. First, for $C_0 = [N]^3$ we have $\phi(C_0) = (|T|/|D|)N^{3\eta} \geq 2^{-d}N^{3\eta}$, and for C we have $\phi(C) = (|T \cap C|/|D \cap C|)|C|^{\eta}$. Since C maximizes $\phi(\cdot)$ we can already make two deductions: first, since $|C| \leq N^3$ we must have $|T \cap C| \geq 2^{-d}|D \cap C|$; and second, since $|T \cap C| \leq |D \cap C|$ we have $|C|^{\eta} \geq 2^{-d}N^{3\eta}$, which implies $|C| \geq 2^{-d/\eta}N^3$. Note that our assumption that $\gamma \leq 2^{-O(dr/\varepsilon)}$ implies that D is γ-pseudorandom with respect to C.

Density. We prove that $\mathbb{E}[h] \geq 2^{-(d+1)}$. Apply Claim 5.2 to T and the cube C. We get

$$\mathbb{E}[h] = \frac{|T \cap C|}{|D \cap C|} \pm O(\gamma^{1/3}) \geq 2^{-d} \pm O(\gamma^{1/3}).$$

The claim follows since $\gamma \leq 2^{-O(d)}$.

Spreadness. We next show that h is (r, ε)-spread. Assume towards a contradiction that there exists a rectangle $R' = X' \times Y' \subset X \times Y$ of size $|R'| \geq 2^{-r}|X||Y|$ such that

$$\mathop{\mathbb{E}}_{(x,y) \in X' \times Y'} h(x,y) > (1+\varepsilon) \mathop{\mathbb{E}}_{(x,y) \in X \times Y} h(x,y).$$

Define $C' = X' \times Y' \times Z$. Since we take γ small enough (concretely, $\gamma \leq 2^{-(d/\eta + r)}$), D is also γ-pseudorandom with respect to C'. We will show that $\phi(C') > \phi(C)$, which contradicts the assumption that C maximizes $\phi(\cdot)$. Applying Claim 5.2 to C and C' gives

$$\mathop{\mathbb{E}}_{(x,y) \in X \times Y} h(x,y) = \frac{|T \cap C|}{|D \cap C|} \pm O(\gamma^{1/3})$$

and

$$\mathop{\mathbb{E}}_{(x,y) \in X' \times Y'} h(x,y) = \frac{|T \cap C'|}{|D \cap C'|} \pm O(\gamma^{1/3})$$

As we have $\gamma \leq (2^{-d}\varepsilon)^{O(1)}$, we get that

$$\frac{|T \cap C'|}{|D \cap C'|} \geq (1 + \varepsilon/2) \frac{|T \cap C|}{|D \cap C|}$$

which gives

$$\frac{\phi(C')}{\phi(C)} \geq (1 + \varepsilon/2) \left(\frac{|C'|}{|C|} \right)^{\eta} \geq (1 + \varepsilon/2) 2^{-r\eta} > 1$$

for $\eta = O(\varepsilon/r)$ small enough.

Mostly left lower-bounded. We next show that h is mostly left lower-bounded. Assume towards a contradiction that there exists $X' \subset X$ of size $|X'| = \beta|X|$ such that

$$\mathop{\mathbb{E}}_{y \in Y} h(x,y) < (1 - \varepsilon) \mathbb{E}[h] \qquad \forall x \in X'.$$

Set $C' = X' \times Y \times Z$. Since we assume that γ is small enough (concretely, $\gamma \leq 2^{-d/\eta}\beta$), D is γ-pseudorandom with respect to C'. Applying Claim 5.2 to C' gives

$$\mathop{\mathbb{E}}_{(x,y) \in X' \times Y} h(x,y) = \frac{|T \cap C'|}{|D \cap C'|} \pm O(\gamma^{1/3}).$$

Repeating the same argument for C, and using the fact that we have $\gamma \leq (2^{-d}\varepsilon)^{O(1)}$, we get

$$\frac{|T \cap C'|}{|D \cap C'|} \leq (1 - \varepsilon) \frac{|T \cap C|}{|D \cap C|} + O(\gamma^{1/3}) \leq (1 - \varepsilon/2) \frac{|T \cap C|}{|D \cap C|},$$

Since D is γ-pseudorandom with respect to C, C', we have $|D \cap C| = (1 \pm \gamma)\mu|C|$, $|D \cap C'| = (1 \pm \gamma)\mu|C'|$ and hence

$$\frac{|T \cap C'|}{|C'|} \leq (1 - \varepsilon/4) \frac{|T \cap C|}{|C|}.$$

Let $C'' = C \setminus C' = (X \setminus X') \times Y \times Z$. We will show that $\phi(C'') > \phi(C)$, which is a contradiction to the maximality of C. Note that

$$|T \cap C''| = |T \cap C| - |T \cap C'| \geq (1 - (1 - \varepsilon/4)\beta)|T \cap C|$$

and

$$|D \cap C''| = |D \cap C| - |D \cap C'| = (1 - \beta \pm \gamma)|D \cap C|$$

and

$$|C''| = (1 - \beta)|C|.$$

Thus

$$\frac{\phi(C'')}{\phi(C)} = \frac{|T \cap C''|}{|T \cap C|}\frac{|D \cap C|}{|D \cap C''|}\left(\frac{|C''|}{|C|}\right)^{\eta} \geq (1 - \beta + \varepsilon\beta/4)(1 - \beta + \gamma)^{-1}(1 - \beta)^{\eta} > 1$$

where the last inequality holds for $\gamma \leq O(\beta\varepsilon), \eta \leq O(\varepsilon)$ small enough. □

As we discussed, we prove Lemma 2.13 by pruning the cube obtained by Lemma 5.5.

Proof of Lemma 2.13. Apply Lemma 5.5 with parameters $d, r+1, \varepsilon/2, \beta = O(2^{-d}\varepsilon)$, which we can as we assume $\gamma = 2^{-\Omega(d^3)}$. Let $C = X \times Y \times Z$ and $f : X \times Z \to [0,1]$, $g : Y \times Z \to [0,1]$, $h : X \times Y \to [0,1]$ be the obtained functions, satisfying the condition that $T \cap C$ viewed as a subset of C is $(d+1, r+1, \varepsilon/2, \beta)$-well behaved. We next prune C to obtain the desired cube and corresponding functions.

Let $X' \subset X$ be the set of points x where f is $(\varepsilon/2)$-left lower-bounded, and $Y' \subset Y$ be the set of points y where g is $(\varepsilon/2)$-left lower-bounded, both with respect to C. Let $C' = X' \times Y' \times Z$ and let f', g', h' be the restrictions of f, g, h to $X' \times Z, Y' \times Z, X' \times Y'$, respectively. We claim that $T \cap C'$, viewed as a subset of the cube C', is $(d+2, r, \varepsilon)$ well-behaved, witnessed by f', g', h'.

We first show that $\mathbb{E}[f'], \mathbb{E}[g'], \mathbb{E}[h'] \geq 2^{-(d+2)}$. We show this for f', and an analogous argument works for g', h'. Note that since $|X'| \geq (1-\beta)|X|$ and f takes values in $[0,1]$, we have

$$\mathbb{E}[f'] = \mathbb{E}[f] \pm O(\beta).$$

Since we know $\mathbb{E}[f] \geq 2^{-(d+1)}$, taking $\beta = O(2^{-d})$ small enough guarantees that $\mathbb{E}[f'] \geq 2^{-(d+2)}$.

We next show that f', g' are (r, ε)-spread. We show this for f', and an analogous argument works for g'. Assume that $R \subset X' \times Z$ is a rectangle of size $|R| \geq 2^{-r}|X'||Z|$. We can also view R as a rectangle $R \subset X \times Z$ of size $|R| \geq (1-\beta)2^{-r}|X||Z| \geq 2^{-(r+1)}|X||Z|$. Recalling that f' is a restriction of f, and applying the assumption that f is $(r+1, \varepsilon/2)$-spread gives

$$\mathop{\mathbb{E}}_{(x,z)\in X'\times Z} f'(x,z) = \mathop{\mathbb{E}}_{(x,z)\in X'\times Z} f(x,z) \leq (1+\varepsilon/2)\,\mathbb{E}[f] \leq (1+\varepsilon)\,\mathbb{E}[f'],$$

where in the last inequality we use the fact that $\mathbb{E}[f'] = \mathbb{E}[f] \pm O(\beta)$ and our choice of $\beta = O(2^{-d}\varepsilon)$.

Finally, we show that f', g' are ε-left lower-bounded. We show this for f', and an analogous argument works for g'. Take any $x \in X'$. We have by assumption

$$\mathop{\mathbb{E}}_{z\in Z}[f(x,z)] \geq (1-\varepsilon/2)\,\mathbb{E}[f].$$

We already saw that $\mathbb{E}[f'] = \mathbb{E}[f] \pm \beta$, and so for $\beta = O(2^{-d}\varepsilon)$ we get that

$$\mathop{\mathbb{E}}_{z\in Z}[f'(x,z)] = \mathop{\mathbb{E}}_{z\in Z}[f(x,z)] \geq (1-\varepsilon/2)\,\mathbb{E}[f] \geq (1-\varepsilon)\,\mathbb{E}[f'].$$

□

6 Construction of sparse pseudorandom sets

We prove in this section the three lemmas from Section 2.5: Claim 2.15 and Lemmas 2.16 and 2.17.

Proof of Claim 2.15. Let $(x, y, z) \in [N]^3$ denote the inputs to $D = D(\mathrm{Col})$. Each player sees two out of three inputs; namely, the first player sees (y, z), the second (x, z), and the third (x, y). Each player computes the color of its respective edge, namely $c_1 = \mathrm{Col}(y, z)$, $c_2 = \mathrm{Col}(x, z)$, $c_3 = \mathrm{Col}(x, y)$. They then need to decide if $c_1 = c_2 = c_3$. This can be easily done by using a randomized protocol for equality between each pair of players, which requires sending only $O(1)$ bits using public randomness. □

We next show that the inner product function is a good expander-coloring. The proof uses standard arguments based on Fourier analysis.

Proof of Lemma 2.17. Let $\mathbb{F}_q$ be a finite field, $k \geq 3$ and set $\eta = q^{-(k-2)/4}$. Let $X, Y \subset \mathbb{F}_q^k$ of size $|X|, |Y| \geq \eta q^k$. Fix any value $v \in \mathbb{F}_q$. We need to show that

$$\Pr_{(x,y)\in X\times Y}[\langle x, y\rangle = v] = q^{-1}(1 \pm \eta).$$

We prove this using Fourier analysis. The additive characters of $\mathbb{F}_q$ are $\chi_u : \mathbb{F}_q \to \mathbb{C}^*$ for $u \in \mathbb{F}_q$. Given $a \in \mathbb{F}_q$, we have $\mathbb{E}_u[\chi_u(a)] = \mathbb{1}[a = 0]$, and hence

$$\Pr_{(x,y)\in X\times Y}[\langle x, y\rangle = v] = \mathop{\mathbb{E}}_{(x,y)\in X\times Y} \mathop{\mathbb{E}}_{u\in\mathbb{F}_q} \left[\chi_u(\langle x, y\rangle - v)\right].$$

Using the fact that $\chi_0 \equiv 1$, $\chi_u(a + b) = \chi_u(a)\chi_u(b)$ and $|\chi_u(-v)| = 1$ we get

$$\Pr_{(x,y)\in X\times Y}[\langle x, y\rangle = v] - q^{-1} \leq q^{-1} \sum_{u\in\mathbb{F}_q\setminus\{0\}} \left| \mathop{\mathbb{E}}_{(x,y)\in X\times Y} [\chi_u(\langle x, y\rangle)] \right|.$$

To conclude the proof we bound the latter sum using Lindsey's lemma (see e.g. [CG88]). Fix $u \in \mathbb{F}_q \setminus \{0\}$. Let $1_X, 1_Y \in \{0, 1\}^{q^k}$ be the indicator vectors of X, Y, respectively. Let H be the corresponding Fourier transform matrix over $\mathbb{F}_q^k$, namely $H_{x,y} = \chi_u(\langle x, y\rangle)$ for $x, y \in \mathbb{F}_q^k$. It is well known that $HH^* = q^k I$ and hence its spectral norm is $\|H\| = q^{k/2}$. Thus

$$\left| \mathop{\mathbb{E}}_{(x,y)\in X\times Y} [\chi_u(\langle x, y\rangle)] \right| = \frac{|1_X H 1_Y|}{|X||Y|} \leq \frac{\|1_X\|_2 \|1_Y\|_2 \|H\|}{|X||Y|} = \frac{\|H\|}{\sqrt{|X||Y|}} \leq \frac{1}{\eta q^{k/2}} = q^{-1}\eta,$$

where the last equality follows by our choice of η. □

We now move to prove Lemma 2.16. We first develop counting lemmas for expanders, which we then apply to prove the lemma.

6.1 Counting lemma for bi-partite expanders

As a starting point, we develop counting lemmas for a single color class. These effectively are bi-partite expanders, but in a slightly non-standard regime, so we formally define them.

Let $G = (U, V, E)$ be a bi-partite graph with parts of equal size $|U| = |V| = N$. Given sets $X \subset U, Y \subset V$, we denote by $e_G(X, Y)$ the number of edges between X, Y.

Definition 6.1 (Bi-partite expander)**.** *Let $G = (U, V, E)$ be a bi-partite graph with $|U| = |V| = N$. We say that G is a (N, p, η)-expander, if for any sets $X \subset U, Y \subset V$ of size $|X|, |Y| \geq \eta N$ it holds*

$$e_G(X, Y) = p|X||Y|(1 \pm \eta).$$

Next, we extend the definition to k-partite graphs. Let $k \geq 2$, and let $G = (V_1, \ldots, V_k; E)$ be a k-partite graph with parts $V_1, \ldots, V_k$, each of size N. For $i \neq j$ we denote by G_{ij} the induced bi-partite graph between parts V_i, V_j, and shorthand $E_{ij}(G) = E(G_{ij})$.

Definition 6.2 (k-partite expander)**.** *Let G be a k-partite graph, with parts each of size N. We say that G is a k-partite (N, p, η)-expander if for any $i \neq j$, the bi-partite graph G_{ij} is an (N, p, η)-expander.*

A graph H on k nodes is said to be a *labeled graph* if its nodes are labeled by $1, \ldots, k$.

Definition 6.3 (Labeled graph homomorphism)**.** *Let G be a k-partite (N, p, η)-expander with parts $V_1, \ldots, V_k$. Let H be labeled graph on k nodes. A tuple $(v_1, \ldots, v_k) \in V_1 \times \cdots \times V_k$ is a homomorphism from H to G if edges of H map to edges of G; that is, if it satisfies*

$$\forall (i, j) : \ (i, j) \in E(H) \Rightarrow (v_i, v_j) \in E_{ij}(G).$$

We denote by $Hom(H, G)$ the set of all such tuples.

Our main goal in this section is to prove the following counting lemma.

Lemma 6.4. *Let G be a k-partite (N, p, η)-expander, let H a labeled graph with k nodes and ℓ edges, and let $U_i \subset V_i$ for $i \in [k]$. Then the number of homomorphisms $(v_1, \ldots, v_k)$ from H to G, that satisfy $v_i \in U_i$ for all i, satisfies*

$$|Hom(H, G) \cap (U_1 \times \cdots \times U_k)| = p^\ell \prod_{i \in [k]} |U_i| \pm 6\eta \ell N^k.$$

Before proving Lemma 6.4, we need the following claim.

Claim 6.5. *Let G be a (N, p, η)-expander with parts U, V. Let $X \subset U, Y \subset V$ of size $|X|, |Y| \geq \eta N$. Define*

$$X' = \{x \in X : e_G(\{x\}, Y) = p|Y|(1 \pm \eta)\}.$$

Then $|X \setminus X'| \leq 2\eta N$.

Proof. Define

$$X_1 = \{x \in X : e_G(\{x\}, Y) > p|Y|(1 + \eta)\}, \qquad X_2 = \{x \in X : e_G(\{x\}, Y) < p|Y|(1 - \eta)\}.$$

Since $|Y| \geq \eta N$ we have $|X_1|, |X_2| < \eta N$. The claim follows since $X \setminus X' \subset X_1 \cup X_2$. □

Proof of Lemma 6.4. We shorthand $\mathcal{H} = \mathrm{Hom}(H, G) \cap (U_1 \times \cdots \times U_k)$. The proof is by induction on k. We start with some base cases. If H has no edges then clearly $|\mathcal{H}| = \prod_{i \in [k]} |U_i|$ and the lemma holds. Similarly, if H has an isolated node then the claim easily reduces to $k - 1$ by removing this node from H, and the corresponding part from G, so may assume H has no isolated nodes. Finally, note that if some set U_i has size $|U_i| \leq \eta N$, then $|\mathcal{H}| \leq \prod |U_i| \leq \eta N^k$, and the lemma also holds. Thus, we may assume that $|U_i| \geq \eta N$ for all $i \in [k]$.

The base case of the induction is $k = 2$, where H consists of a single edge $(1, 2)$. In this case, by definition of a (N, p, η)-expander we have

$$|\mathcal{H}| = |\mathrm{Hom}(H, G) \cap (U_1 \times U_2)| = e_G(U_1, U_2) = p|U_1||U_2|(1 \pm \eta) = p|U_1||U_2| \pm \eta N^2$$

and the lemma holds.

We next consider $k \geq 3$. For $v_k \in U_k$, define $\mathcal{H}(v_k) = \{(v_1, \ldots, v_{k-1}) : (v_1, \ldots, v_k) \in \mathcal{H}\}$. Then

$$|\mathcal{H}| = \sum_{v_k \in U_k} |\mathcal{H}(v_k)|.$$

Assume the node $k \in V(H)$ has s neighbours in H, which we may assume without loss of generality are $1, \ldots, s$. Let $\Gamma_i(v_k) = \{v_i \in U_i : (v_i, v_k) \in E_{ik}(G)\}$ denote the neighbours of v_k in U_i for $i \in [s]$. Let G' be the $(k-1)$-partite graph obtained from G by removing the part V_k. Let H' be the graph obtained from H by removing node k. Observe that G' is a $(k-1)$-partite (N, p, η)-expander, that H' is a labeled graph with $k - 1$ nodes and $\ell - s$ edges, and that

$$\mathcal{H}(v_k) = \mathrm{Hom}(H', G') \cap (\Gamma_1(v_k) \times \cdots \times \Gamma_s(v_k) \times U_{s+1} \times \cdots \times U_{k-1}).$$

Using the induction hypothesis for G', H' gives

$$|\mathcal{H}(v_k)| = p^{\ell-s} \prod_{i=1}^{s} |\Gamma_i(v_k)| \cdot \prod_{i=s+1}^{k-1} |U_i| \pm 6\eta(\ell - s)N^{k-1}.$$

Next, define

$$U'_k = \{v_k \in U_k : \forall i \in [s],\ |\Gamma_i(v_k)| = p|U_i|(1 \pm \eta)\}.$$

Applying Claim 6.5, we have $|U_k \setminus U'_k| \leq 2s\eta N$. We now complete the calculations. Note that we may assume $\eta \leq 1/6\ell$ otherwise the bound is trivial; in this regime we have $(1 \pm \eta)^s = 1 \pm 2\eta s$. For $v_k \in U'_k$ we have

$$|\mathcal{H}(v_k)| = p^{\ell} \prod_{i=1}^{k-1} |U_i|(1 \pm \eta)^s \pm 6(\ell - s)\eta N^{k-1} = p^{\ell} \prod_{i=1}^{k-1} |U_i| \pm (6\ell - 3s)\eta N^{k-1}.$$

For $v_k \in U_k \setminus U'_k$ we naively bound

$$\sum_{v_k \in U_k \setminus U'_k} |\mathcal{H}(v_k)| \leq |U_k \setminus U'_k| N^{k-1} \leq 2s\eta N^k.$$

Summing over all $v_k \in U_k$, we conclude that

$$|\mathcal{H}| = p^{\ell} \prod_{i=1}^{k} |U_i| \pm 6\eta\ell N^{k-1}.$$

□

6.2 Counting lemma for partite expander-colorings

We now apply the counting lemma for expanders (Lemma 6.4) to count monochromatic patterns in expander-colorings. Let $K_N^{(k)}$ denote the complete k-partite graph, with parts $V_1, \ldots, V_k$, each of size N. We identify each V_i with $[N]$ when possible to do so without confusion. We consider edge-colorings of $K_N^{(k)}$. We denote them by $\mathrm{Cols} = (\mathrm{Col}_{ij} : 1 \le i < j \le k)$ where each $\mathrm{Col}_{ij} : [N]^2 \to [M]$.

Definition 6.6 (k-partite expander-coloring). *An edge-coloring Cols of $K_N^{(k)}$ is a k-partite (N, M, η)-expander coloring if Col_{ij} are (N, M, η)-expander colorings for all $i \neq j$.*

Definition 6.7 (Monochromatic patterns). *Let Cols be a k-partite (N, M, η)-expander coloring. Let H be labeled graph on k nodes. A tuple $(v_1, \ldots, v_k) \in [N]^k$ is a monochromatic copy of H in G if all edges of H are mapped to edges with the same color under Cols. Namely, if there exists $m \in [M]$ such that*

$$\forall (i,j) : (i,j) \in E(H) \Rightarrow Col_{ij}(v_i, v_j) = m.$$

We denote by $Mon(H, Cols)$ the set of all such tuples.

The following lemma is an application of Lemma 6.4 to count monochromatic patterns inside partite expander-colorings.

Lemma 6.8. *Let Cols be a k-partite (N, M, η)-expander coloring, let H a labeled graph with k nodes and ℓ edges, and let $U_i \subset V_i$ for $i \in [k]$. Then*

$$|Mon(H, Cols) \cap (U_1 \times \cdots \times U_k)| = M^{1-\ell} \prod_{i \in [k]} |U_i| \pm 3\eta\ell M N^k.$$

Proof. For each color $m \in [M]$, let G_m be the k-partite graph corresponding to edges in Cols colored in color m. By definition, G_m is a k-partite (N, M^{-1}, η)-expander. The lemma follows by applying Lemma 6.4 to each G_m and summing over all m. □

6.3 Monochromatic triangles in expander-colorings

We prove Lemma 2.16 based on Lemma 6.8. We need one more definition before giving the proof. Given an expander-coloring $\mathrm{Col} : [N] \times [N] \to [M]$, we denote by $\mathrm{Cols}(\mathrm{Col}, k)$ the k-partite expander-coloring where Col is used as the edge-coloring between each of two parts; namely $\mathrm{Cols}(\mathrm{Col}, k) = (\mathrm{Col}_{ij} = \mathrm{Col} : 1 \le i < j \le k)$.

Proof of Lemma 2.16. Let $\mathrm{Col} : [N] \times [N] \to [M]$ be an (N, M, η)-expander coloring. Let $D = D(\mathrm{Col})$. We will show that D is γ-pseudorandom if we choose η small enough. Let $C = X \times Y \times Z$ be a cube of size $|C| \ge \gamma N^3$. Define $v : X \times Y \to \mathbb{R}_{\ge 0}$ as

$$v(x, y) = \mathop{\mathbb{E}}_{z \in Z} D(x, y, z).$$

We need to show that $\mathbb{E}[v] = (1 \pm \gamma) M^{-2}$ and $\mathbb{E}[v^2] = (1 \pm \gamma) M^{-4}$.

We first analyze the first moment of v. Define a 3-partite (N, M, η)-expander coloring $\mathrm{Cols}_1 = \mathrm{Cols}(\mathrm{Col}, 3)$. Let H_1 be a triangle. Observe that

$$\mathbb{E}[v] = \frac{|\mathrm{Mon}(H_1, \mathrm{Cols}_1) \cap (X \times Y \times Z)|}{|X||Y||Z|}.$$

Lemma 6.8 then gives

$$\mathbb{E}[v] = M^{-2} \pm O(\eta M N^3/|X||Y||Z|) = M^{-2} \pm O(\eta M/\gamma) = (1 \pm \gamma)M^{-2}$$

since we assume $\eta = O(\gamma^2 M^{-3})$.

Next, we analyze the second moment of v. Define a 4-partite (N, M, η)-expander coloring $\mathrm{Cols}_2 = \mathrm{Cols}(\mathrm{Col}, 4)$. Let H_2 be a union of two triangles sharing an edge; concretely, $V(H_2) = \{1, 2, 3, 4\}$ and $E(H_2) = \{(1, 2), (1, 3), (2, 3), (1, 4), (2, 4)\}$. Observe that

$$\mathbb{E}[v^2] = \frac{|\mathrm{Mon}(H_2, \mathrm{Cols}_2) \cap (X \times Y \times Z \times Z)|}{|X||Y||Z|^2}.$$

Lemma 6.8 then gives

$$\mathbb{E}[v^2] = M^{-4} \pm O(\eta M N^4/|X||Y||Z|^2) = M^{-4} \pm O(\eta M/\gamma^2) = (1 \pm \gamma)M^{-4}$$

since we assume $\eta = O(\gamma^3 M^{-5})$. □

7 Open problems

We view Theorem 2.8 as the main new technical innovation of this work, and Theorem 1.1 as an application of it. One natural open problem is to extend the proof to other functions, and in particular to ExactlyN; a strong lower bound for the NOF deterministic complexity of it would imply strong lower bounds for the corners problem. Another open problem is to extend the proof to more than 3 players. The challenges in it appear to be similar to those of extending the results of [KM23] from three-term arithmetic progressions to longer progressions.

References

[AB23] Josh Alman and Jarosław Błasiok. Matrix multiplication and number on the forehead communication. *arXiv preprint arXiv:2302.11476*, 2023.

[AS20] Noga Alon and Adi Shraibman. Number on the forehead protocols yielding dense Ruzsa–Szemerédi graphs and hypergraphs. *Acta Mathematica Hungarica*, 161(2):488–506, 2020.

[BDPW10] Paul Beame, Matei David, Toniann Pitassi, and Philipp Woelfel. Separating deterministic from randomized multiparty communication complexity. *Theory of Computing*, 6(1):201–225, 2010.

[BGG06] Richard Beigel, William Gasarch, and James Glenn. The multiparty communication complexity of Exact-T: Improved bounds and new problems. In *International Symposium on Mathematical Foundations of Computer Science*, pages 146–156. Springer, 2006.

[BH12] Paul Beame and Trinh Huynh. Multiparty communication complexity and threshold circuit size of AC^0. *SIAM Journal on Computing*, 41(3):484–518, 2012.

[BNS89] László Babai, Noam Nisan, and Mario Szegedy. Multiparty protocols and logspace-hard pseudorandom sequences. In *Proceedings of the twenty-first annual ACM symposium on Theory of computing*, pages 1–11, 1989.

[BPS07] Paul Beame, Toniann Pitassi, and Nathan Segerlind. Lower bounds for Lovász–Schrijver systems and beyond follow from multiparty communication complexity. *SIAM Journal on Computing*, 37(3):845–869, 2007.

[BPSW06] Paul Beame, Toniann Pitassi, Nathan Segerlind, and Avi Wigderson. A strong direct product theorem for corruption and the multiparty communication complexity of disjointness. *computational complexity*, 15:391–432, 2006.

[CFL83] Ashok K Chandra, Merrick L Furst, and Richard J Lipton. Multi-party protocols. In *Proceedings of the fifteenth annual ACM symposium on Theory of computing*, pages 94–99, 1983.

[CG88] Benny Chor and Oded Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM Journal on Computing*, 17(2):230–261, 1988.

[CP10] Arkadev Chattopadhyay and Toniann Pitassi. The story of set disjointness. *ACM SIGACT News*, 41(3):59–85, 2010.

[DPV09] Matei David, Toniann Pitassi, and Emanuele Viola. Improved separations between nondeterministic and randomized multiparty communication. *ACM Transactions on Computation Theory (TOCT)*, 1(2):1–20, 2009.

[Gow01] William T Gowers. A new proof of Szemerédi's theorem. *Geometric & Functional Analysis (GAFA)*, 11(3):465–588, 2001.

[Gow06] W Timothy Gowers. Quasirandomness, counting and regularity for 3-uniform hypergraphs. *Combinatorics, Probability and Computing*, 15(1-2):143–184, 2006.

[Hat10] Hamed Hatami. Graph norms and Sidorenko's conjecture. *Israel Journal of Mathematics*, 175:125–150, 2010.

[KM23] Zander Kelley and Raghu Meka. Strong bounds for 3-progressions. *arXiv preprint arXiv:2302.05537*, 2023.

[KN96] Eyal Kushilevitz and Noam Nisan. Communication complexity, 1996.

[LLSW16] Troy Lee, Nikos Leonardos, Michael Saks, and Fengming Wang. Hellinger volume and number-on-the-forehead communication complexity. *Journal of Computer and System Sciences*, 82(6):1064–1074, 2016.

[LPS18] Nati Linial, Toniann Pitassi, and Adi Shraibman. On the communication complexity of high-dimensional permutations. In *10th Innovations in Theoretical Computer Science Conference (ITCS 2019)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.

[LS09] Troy Lee and Adi Shraibman. Disjointness is hard in the multiparty number-on-the-forehead model. *Computational Complexity*, 18:309–336, 2009.

[LS21a] Nati Linial and Adi Shraibman. An improved protocol for the exactly-n problem. In *36th Computational Complexity Conference (CCC 2021)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2021.

[LS21b] Nati Linial and Adi Shraibman. Larger corner-free sets from better NOF exactly-n protocols. *Discrete Analysis*, 10 2021.

[Raz00] Ran Raz. The BNS-Chung criterion for multi-party communication complexity. *Computational Complexity*, 9:113–122, 2000.

[RY20] Anup Rao and Amir Yehudayoff. *Communication Complexity: and Applications*. Cambridge University Press, 2020.

[Shr18] Adi Shraibman. A note on multiparty communication complexity and the Hales–Jewett theorem. *Information Processing Letters*, 139:44–48, 2018.