Electronic Colloquium on Computational Complexity, Report No. 170 (2024)

Corners in Quasirandom Groups via Sparse Mixing

Michael Jaber* Shachar Lovett† Anthony Ostuni‡

Abstract

We improve the best known upper bounds on the density of corner-free sets over quasirandom groups from inverse poly-logarithmic to quasi-polynomial. We make similarly substantial improvements to the best known lower bounds on the communication complexity of a large class of permutation functions in the 3-player Number-on-Forehead model. Underpinning both results is a general combinatorial theorem that extends the recent work of Kelley, Lovett, and Meka (STOC'24), itself a development of ideas from the breakthrough result of Kelley and Meka on three-term arithmetic progressions (FOCS'23).

1 Introduction

In the early 1980s, Chandra, Furst, and Lipton introduced the Number-on-Forehead (NOF) model of communication complexity [CFL83] to better capture interaction with shared information. The k-NOF model is defined by k players communicating over a shared channel in order to compute a function $f : (\{0,1\}^n)^k \to \{0,1\}$. Each player can see the $k-1$ inputs of every other player, but they cannot see their own. Albeit perhaps unintuitive at first glance, the model has a number of strikingly powerful and surprising connections to other areas of theoretical computer science and combinatorics. For example, lower bounds for $k = \omega(\log n)$ players would imply breakthrough circuit lower bounds [BNS89, NW91, Raz00, BH12], and the communication complexity of several natural functions is known to be equivalent to central problems in Ramsey theory [CFL83, Shr18, LS21b].

Unfortunately, our understanding of this model is severely lacking. Only in the past year have researchers discovered explicit functions witnessing strong separations between randomized and deterministic 3-NOF communication complexity [KLM24], despite the fact that optimal separations were long known to exist non-explicitly [BDPW10]. More precisely, Kelley, Lovett, and Meka exhibited an explicit 3-player function which has a constant cost randomized protocol, but requires $\Omega(n^{1/3})$ bits of communication to compute deterministically. Their primary technical tool is a combinatorial adaptation of ideas from the recent breakthrough of Kelley and Meka on three-term arithmetic progressions (3APs) [KM23].

One of the most well-studied functions in the NOF setting is $\mathsf{Exactly\text{-}N}$, where each player receives a number in $[N] := \{1, 2, \ldots, N\}$, and they wish to determine if their numbers sum to N. Introduced by [CFL83], they showed that the 3-NOF complexity of $\mathsf{Exactly\text{-}N}$ is at most $O(\sqrt{\log N})$ using the Behrend construction of 3-AP free sets [Beh46]. In fact, they observed a near equivalence between $\mathsf{Exactly\text{-}N}$ (for three players) and the size of sets $S \subset [N]^2$ without *corners*: three points $(x, y), (x + z, y), (x, y + z) \in [N]^2$ with $z \neq 0$. The first nontrivial bounds on the size of such sets were proven earlier by Ajtai and Szemerédi [AS74], but the quantitative behavior was poor, since the proof relied on Szemerédi's regularity lemma [Sze75]. The strongest

*Department of Computer Science, UT Austin. Email: mjjaber@cs.utexas.edu. Supported by a Simons Investigator Award (#409864, David Zuckerman).

†Department of Computer Science and Engineering, UC San Diego. Email: slovett@ucsd.edu. Supported by a Simons Investigator Award (#929894).

‡Department of Computer Science and Engineering, UC San Diego. Email: aostuni@ucsd.edu. Supported by a Simons Investigator Award (#929894, Shachar Lovett).

bound to date is due to Shkredov [Shk06], who showed any corner-free set of size δN^2 must satisfy

$$\delta \leqslant O\left(\frac{1}{(\log\log N)^c}\right)$$

for some absolute constant $c > 0$ (see also [Shk05] and the exposition by Green over finite fields [Gre04, Gre05]). Using the connection shown by [CFL83], this implies that the complexity of Exactly-N is at least $\Omega(\log\log\log N)$. Unfortunately, the techniques developed in [KLM24] do not suffice to improve this lower bound, since they only apply to much denser functions. Concretely, the number of solutions of Exactly-N (that is, triples $x, y, z \in [N]$ that satisfy $x + y + z = N$) is at most N^2, whereas the function exhibited by [KLM24] to have strong lower bounds for deterministic NOF communication has N^{3-c} solutions for some small constant $c > 0$; their technique is restricted to such functions.

Observe that corners may be viewed as a multidimensional generalization of arithmetic progressions. In fact, upper bounds on corner-free sets easily imply upper bounds on 3AP-free sets (see e.g. [Zha23, Section 2.4]). Given their tight relationship, many researchers have suspected that the recent techniques of Kelley and Meka [KM23] used to improve bounds for 3AP-free sets will be amenable to usage in the case of corners (see e.g. [Mek23] and [Pel23, Section 1.2]). While there is some preliminary evidence that this direction is viable [JLO24, Mil24], such strong bounds remain currently beyond reach.

1.1 Our results

A common strategy in additive combinatorics when working over the integers is to prove a similar result in some model setting, such as finite fields, then port the result back to the integers using standard machinery. One interesting setting is *quasirandom groups*. For now, one can think of a quasirandom group as a finite group G enjoying the property that any two large sets $A, B \subset G$ "mix" under convolutions. In other words, if we take random samples $a \in A$ and $b \in B$, then the distribution of ab is close to the uniform distribution over G. A classic example of a quasirandom group is $G = \mathrm{SL}_2(\mathbb{F}_p)$, the set of 2×2 matrices over the finite field $\mathbb{F}_p$ for p prime with determinant 1. The Exactly-N problem naturally generalizes to any finite group G [BGG06], where the players receive inputs $x, y, z \in G$ and accept if and only if their inputs satisfy $xyz = 1_G$. Note that over any group, Exactly-N has a constant cost randomized protocol by reducing to equality. We obtain the following lower bound for computing Exactly-N over $G = \mathrm{SL}_2(\mathbb{F}_p)$ by either deterministic or even non-deterministic protocols.

Theorem 1.1 (Special case of Theorem 4.3)**.** *Any non-deterministic 3-NOF protocol computing* Exactly-N over $G = \mathrm{SL}_2(\mathbb{F}_p)$ for prime p requires $\Omega(\log^{1/4}|G|)$ bits of communication.

Similar to the abelian case, there exists an intimate connection between Exactly-N over a group G and corner-free sets in $G \times G$. However, there is a slight subtlety here, as corners generalize to the non-abelian setting in two non-equivalent ways. One option is triples of the form $\{(x, y), (zx, y), (x, zy)\}$ for $z \neq 1_G$, often referred to as *naïve corners*. In this setting, Austin [Aus16] proved that for $G = \mathrm{SL}_2(\mathbb{F}_p)$, any subset of $G \times G$ without naïve corners has size $|G|^{2-\varepsilon}$ for some small constant $\varepsilon > 0$. Alternatively, one can consider triples of the form $\{(x, y), (xz, y), (x, zy)\}$ for $z \neq 1_G$, sometimes called *BMZ corners* after the first researchers to study them [BMZ97]. This formulation is less understood, and it corresponds to the three-player Exactly-N function over general groups (see e.g. [Vio19, Lemma 21]). We will focus our attention on this latter generalization, and henceforth refer to them simply as corners. Austin also showed that corner-free sets over $\mathrm{SL}_2(\mathbb{F}_p)$ have density at most $\delta \leqslant O(1/\log^c|G|)$ for some absolute constant $c > 0$ (see [Vio19, Section 5] for a nice exposition). We are able to substantially improve this bound.

Theorem 1.2 (Special case of Corollary 4.10)**.** *Let* $G = \mathrm{SL}_2(\mathbb{F}_p)$ *for prime* p*. Then, any corner-free subset of* $G \times G$ *has size at most* $\delta|G|^2$ *for*

$$\delta \leqslant \exp\left(-\Omega\left(\log^{1/4}|G|\right)\right).$$

We emphasize that Theorems 1.1 and 1.2 are only special cases of more general theorems, and we direct readers to Sections 4.1 and 4.2, respectively, for details.

Both of our results are consequences of a general combinatorial theorem which may be of independent interest. Before stating it, we require some definitions. A set $S \subset [N]^3$ is called a *permutation function*[1] if for any fixing of two coordinates of some $a \in [N]^3$, there is precisely one choice of the other coordinate such that $a \in S$. For example, $S = \{(x, y, z) : xyz = 1_G\} \subset G^3$ is an example of a permutation function (identifying $[N]$ with G). Given a permutation function $S \subset [N]^3$ and a subset $A \subset S$, we denote by $A_{XY}, A_{XZ}, A_{YZ} \subset [N]^2$ the projections of A to the XY, XZ, YZ-faces of $[N]^3$, respectively.

Theorem 1.3 (Informal special case of Theorem 2.4)**.** *Let $d \geqslant 1$. Suppose $S \subset [N]^3$ is a permutation function and $A \subset S$ is a set of size $|A| \geqslant 2^{-d}|S|$. If S is sufficiently pseudorandom (in the sense of Definition 2.1), then*

$$\left|\left\{(x, y, z) \in [N]^3 : (x, y) \in A_{XY}, (x, z) \in A_{XZ}, (y, z) \in A_{YZ}\right\}\right| \geqslant 2^{-O(d^3)} N^3.$$

For now, the reader can think of the pseudorandomness condition as saying that the density of S stays roughly the same whenever you restrict to some large cube. Theorem 1.3 should be compared with [KLM24, Lemma 2.10], where they refer to the quantity on the left-hand side of the above inequality as the "cylinder intersection closure of A." The two results can be viewed as similar statements in two extreme regimes for the set S. Our theorem holds when S is permutation function, so it must necessarily be sparse (of size $|S| = N^2$), whereas their result holds in the dense case where S has size roughly N^{3-c} for some small enough constant $c > 0$. We briefly note that our pseudorandomness notion differs from theirs to better reflect an alternative regime of interest.

1.2 Future work

We conclude by noting a few directions for future work. The results of [KLM24] hold for sufficiently dense functions, while our results apply only to permutation functions which are sparse. It would be interesting to see if these results can be unified in a theorem which works in all density regimes. Another natural open question is to extend Theorem 1.2 to give quasi-polynomial bounds for corner-free sets over the integers or $\mathbb{F}_2^n$. Over the integers, there are constructions of corner-free sets of size $2^{-\Omega(\sqrt{\log N})} N^2$ [Beh46] (see also the recent improvements [LS21a, Gre21, Hun22]). Thus, such an extension would be optimal in the "shape" of the bound. While we are optimistic that the techniques present here may be useful in these settings, we are not able to directly apply Theorem 1.3, since the corresponding ambient set (see Section 4.2 for more details)

$$S := \{(x, y, x + y) \in (\mathbb{Z}/N\mathbb{Z})^3 : x, y \in \mathbb{Z}/N\mathbb{Z}\}$$

is not sufficiently pseudorandom (in the sense of Definition 2.1). For instance, if $X, Y = \{1, \ldots, N/4\}$ and $Z = \{3N/4, \ldots, N - 1\}$, then the cube $X \times Y \times Z$ is dense in $(\mathbb{Z}/N\mathbb{Z})^3$ but contains no points in S. Similar obstructions also exist if we replace $\mathbb{Z}/N\mathbb{Z}$ with other abelian groups.

Along similar lines, we note the bound in Theorem 1.2 appears to essentially be the quantitative limit of our techniques. However, it remains plausible that the strong structure imbued by quasirandomness guarantees that the largest corner-free sets over $G = \mathrm{SL}_2(\mathbb{F}_p)$ have size $|G|^{2-\varepsilon}$ for some small constant $\varepsilon > 0$. Such bounds would imply optimal separations between randomized and deterministic 3-NOF protocols. It would also be interesting to extend our NOF lower bounds to more than 3 players.

Paper organization. We provide a detailed proof overview of our main theorem in Section 2 with proofs of the main technical lemmas deferred to Sections 5, 6, and 7. Section 3 contains a review of preliminary definitions and facts. Section 4 contains applications to lower bounds in the NOF model of communication, corners in quasirandom groups, and insights about the triangle removal lemma, respectively.

[1] Such sets are called *2-dimensional permutations* in [LL14, LPS19], and extend *graph functions* studied in [BDPW10, Shr18].

Acknowledgments. We thank Russell Impagliazzo and David Zuckerman for helpful conversations, and Ilya Shkredov for answering a question about the current state-of-the-art. MJ would like to thank Sarah Peluse for her encouragement and collaboration on this question, as well as Amey Bhangale and Surya Teja Gavva for their collaboration on this question at the Simons Institue for the Theory of Computing.

2 Proof overview

In this section, we provide a proof overview of our main theorem. Throughout, we recall various definitions and results for the reader's convenience. We begin with a pseudorandomness notion which will be key to our proof.

Definition 2.1 (Pseudorandom against cubes). *A set $S \subset X \times Y \times Z$ is* γ-pseudorandom against cubes *if for every cube $C \subset X \times Y \times Z$, we have*

$$\left| \mathop{\mathbb{E}}_{(x,y,z)\in S}[C(x,y,z)] - \mathop{\mathbb{E}}_{x\in X, y\in Y, z\in Z}[C(x,y,z)] \right| \leqslant \gamma.$$

In other words, the density of a set S which is pseudorandom against cubes cannot change drastically by restricting to a large cube C. Throughout the proof, we will work with $S \subset [N]^3$ which is pseudorandom against cubes. We will also mandate that S is very sparse; in particular, if we fix any two coordinates, there is at most one choice for the last coordinate which produces an element of S. More formally, we define the notion of a (weak) permutation function:

Definition 2.2 ((Weak) permutation function). *Let $f : [N]^3 \to \{0,1\}$. We call f a* permutation function *if for every fixing of any two coordinates, there is exactly one fixing of the remaining coordinate so that $f(x,y,z) = 1$. If instead there exists at most one value, we call f a* weak permutation function*.*

Mandating that S is a (weak) permutation function is important for many of the applications that we give. Indeed, it is one of the main challenges in extending [KLM24] where many of the results only apply when S is relatively dense in $[N]^3$.

We will also want to project a given set $A \subset [N]^3$ to the faces of a given cube C in order to work with two-dimensional sets. We notate the marginals of a set A with respect to a cube C in the following way:

Notation 2.3. *For a set $A \subset [N]^3$ and a cube $C = X \times Y \times Z$, let $A_{XY} \subset X \times Y$ denote the projection of A onto its XY-face. More formally, a point $(x,y) \in A_{XY}$ if and only if there exists $z \in Z$ with $(x,y,z) \in A \cap C$. The sets A_{XZ} and A_{YZ} are defined analogously.*

We state our main result below.

Theorem 2.4. *Let $d, s \geqslant 1$. Suppose $S \subset [N]^3$ is a weak permutation function of size $|S| \geqslant 2^{-s}N^2$ which is γ-pseudorandom against cubes, and $A \subset S$ a set of size $|A| \geqslant 2^{-d}|S|$. For $\gamma \leqslant 2^{-O(d^4+ds)}$ small enough, we have*

$$\left|\left\{(x,y,z) \in [N]^3 : (x,y) \in A_{XY}, (x,z) \in A_{XZ}, (y,z) \in A_{YZ}\right\}\right| \geqslant 2^{-O(d^3+s)}N^3.$$

Before we begin with the proof, we emphasize that our contribution is mostly quantitative. The overall structure of our proof has been present in the literature studying corners for some time (e.g. see [LM05, Shk05] and [Gre04] for an exposition). For example, much of what we will see below when working with respect to a pseudorandom set was present in [LM05], albeit in an arithmetic setting. All of the listed prior work used the standard "box norm" to understand rectangular structure, whereas we give an improvement by working with a higher order variant, known as grid norms. Most of the work in proving Theorem 2.4 goes into pinning down stronger quantitative claims when working with grid norms as opposed to box norms.

The argument proceeds in three main steps. First, we will restrict A to a large cube C where $A \cap C$ satisfies various combinatorial pseudorandom properties. Then, we will show how to efficiently convert these combinatorial statements to analytic conditions. Finally, we will argue that these analytic conditions are enough to imply mixing.

2.1 Obtaining spreadness

The first step in the argument is to restrict A to a large cube C where $A \cap C$ satisfies certain pseudorandom properties. We will want to ensure two pseudorandom properties of the marginals of A to faces of C. First, we want the marginals of A to avoid any strong rectangular structure. We will also want to avoid any rows which are too sparse. This motivates the following definitions of *spreadness* and *left lower-boundedness*, which were introduced in [KLM24].

Definition 2.5 (Spread). *Let $r \geqslant 1$ and $\varepsilon \in (0,1)$. A function $f : X \times Y \to [0,1]$ is* (r,ε)-spread *if for any rectangle $R = X' \times Y' \subset X \times Y$ of size $|R| \geqslant 2^{-r}|X||Y|$, we have*

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

In other words, a function which is spread admits no significant density increment when restricting to a large rectangle.

Definition 2.6 (Left lower-bounded). *Let $\varepsilon \in (0,1)$. A function $f : X \times Y \to [0,1]$ is* ε-left lower-bounded *if for every $x \in X$, we have*

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

Left lower-boundedness ensures that the rows of f are not too sparse. Traditionally, one finds the desired cube C by an iterative process, where if one of the pseudorandom properties is violated, we can find a slightly smaller cube on which the density of our set has increased. The process usually concludes by arguing that the density can only increase finitely many times.

Here is a natural first attempt. Suppose A is a subset of a weak permutation function $S \subset X \times Y \times Z$, where we initially set $X = Y = Z = [N]$. For now, we will *not* use the property that S is pseudorandom against cubes. If the XY-marginal of A is not (r,ε)-spread, then there exists a rectangle $R = X' \times Y' \subset X \times Y$ of size $|R| \geqslant 2^{-r}|X||Y|$ with

$$\mathop{\mathbb{E}}_{x\in X',y\in Y'} A_{XY}(x,y) \geqslant (1+\varepsilon) \mathop{\mathbb{E}}_{x\in X,y\in Y} A_{XY}(x,y).$$

A logical next step is to restrict A to the cube $C = X' \times Y' \times Z$ so that the density of A has increased on the $X'Y'$-face. The issue here is that the marginal of A on some other face of C can decrease significantly. Consider some point $(x,y,z) \in A$. If $y \notin Y'$, then $(x,y,z) \notin A \cap C$. Therefore, the point $(x,z) \in A_{XZ}$ will *not* be in the marginal $A_{X'Z}$. Obtaining a density increment on one marginal might undo progress that was obtained on some other marginal, and it seems the process may never end. This suggests that we need some other measure of progress in our density increment strategy.

To remedy this, [KLM24] exploited the fact that $A \subset S$ where S is pseudorandom against cubes. The pseudorandomness of S implies that for any large cube C,

$$\frac{|S \cap C|}{|C|} \approx \frac{|S|}{N^3}.$$

Thus, the density of S in C will always stay roughly the same across every large cube. This indicates that the density $|A \cap C|/|S \cap C|$ might be a useful measure of progress. We follow the approach used in [KLM24, Lemma 5.5] to obtain a density increment theorem. The proof of the following lemma is deferred to Section 5.

Lemma 2.7 (Restricting to a good cube)**.** *Let $S \subset [N]^3$ be a weak permutation function which is γ-pseudorandom against cubes. Let $d \geqslant 1, r \geqslant 1, \varepsilon \in (0,1)$, and assume $\gamma \leqslant 2^{-\Omega(dr/\varepsilon)}$. Let $A \subset S$ of size $|A| \geqslant 2^{-d}|S|$. Then there is a cube $C = X \times Y \times Z \subset [N]^3$ of size $|C| \geqslant 2^{-O(dr/\varepsilon)}N^3$ with the following properties:*

1. *$|A \cap C| \geqslant 2^{-(d+1)}|S \cap C|$,*
2. *A_{XZ}, A_{YZ} are (r, ε)-spread,*
3. *A_{XZ}, A_{YZ} are ε-left lower-bounded.*

There are two differences worth noting between Lemma 2.7 and [KLM24, Lemma 5.5]. For one, Kelley, Lovett, and Meka work relative to a set $A \subset D$ where the set D is not a (weak) permutation function; in fact, it must be much denser. Concretely, their aim is to apply the theorem with D of size at least N^{3-c} for some small constant $c \in (0,1)$. One challenge that comes with this is that the marginals of D are no longer sets, but rather functions obtained by averaging over a fixed coordinate. For example, the XY-marginal of D is defined by $\mathbb{E}_{z\in Z} D(x,y,z)$. To deal with this, their notion of pseudorandomness against cubes is stronger than ours. It requires that both D is pseudorandom against cubes, and in addition that the marginals of D to faces, when considered inside large cubes, are close to uniform: for a large cube $C = X \times Y \times Z$, the function $\mathbb{E}_{z\in Z} D(x,y,z)$ is close to uniform over $X \times Y$. (Weak) permutation functions cannot satisfy this second property, since the number of $z \in Z$ where $(x,y,z) \in S$ is either 0 or 1. Thus, we have to make some alterations to their proof to get what we need, but the overall ideas are similar.

2.2 Density increment for sparse functions

The next key step in our proof is to convert spreadness into an analytic statement that will let us guarantee mixing. Our main tool for doing this will be the use of grid norms, which were first defined in [KLM24].

Definition 2.8 (Grid norms)**.** *For a function $f : X \times Y \to \mathbb{R}$ and $\ell, k \in \mathbb{N}$, let*

$$\begin{aligned} U_{\ell,k}(f) &= \mathop{\mathbb{E}}_{x_1,\dots,x_\ell \in X} \left(\mathop{\mathbb{E}}_{y\in Y} f(x_1,y)\cdots f(x_\ell,y) \right)^k \\ &= \mathop{\mathbb{E}}_{y_1,\dots,y_k \in Y} \left(\mathop{\mathbb{E}}_{x\in X} f(x,y_1)\cdots f(x,y_k) \right)^\ell \\ &= \mathop{\mathbb{E}}_{\substack{x\in X^\ell \\ y \in Y^k}} \prod_{i=1}^{\ell}\prod_{j=1}^{k} f(x_i,y_j). \end{aligned}$$

The (ℓ,k)-grid norm *of f is given by $\|f\|_{U(\ell,k)} := |U_{\ell,k}(f)|^{1/\ell k}$.*

The purpose of the grid norm is to measure rectangular structure. The reader may notice that the $(2,2)$-grid norm corresponds to the classical "box norm," which has become a staple in studying corners and other additive combinatorial problems. The reader can check that rectangles of density δ can have grid norms much larger than δ, while random sets of density δ have grid norms roughly δ. [KLM24] showed that if the grid norm of f is significantly larger than its expectation, then we can find a large rectangle under which f admits a density increment.

Lemma 2.9 ([KLM24, Lemma 4.7])**.** *Let $f : X \times Y \to [0,1]$; suppose that $\|f\|_1 \geqslant \delta$. Let $\ell, k \in \mathbb{N}$. If*

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

then there exists some rectangle $R \subset X \times Y$ with

$$\mathop{\mathbb{E}}_{(x,y)\in R} f(x,y) \geqslant \left(1 + \frac{\varepsilon}{2}\right)\|f\|_1 \quad \text{and} \quad \|R\|_1 \geqslant \frac{1}{2}\cdot\varepsilon\cdot\delta^{\ell k+1}.$$

Stated in the contrapositive, the above theorem shows that if f is (r, ε)-spread for

$$r \geqslant (\ell k + 1)\log(1/\delta) + \log(1/\varepsilon),$$

then $\|f\|_{U(\ell,k)} \leqslant (1+2\varepsilon)\|f\|_1$. It is worth noting why (ℓ, k)-grid norms give an advantage over the classical $(2,2)$-grid norm. This is a key aspect of our work which allows us to prove stronger bounds. The grid norm arises naturally when trying to control expressions of the form

$$\Lambda(f,g,h) = \mathop{\mathbb{E}}_{x\in X, y\in Y, z\in Z} f(x,y)g(x,z)h(y,z),$$

as is the case in Theorem 2.4 with $f = A_{XY}, g = A_{XZ}, h = A_{YZ}$. A simple application of Hölder's inequality gives the following claim, which is in the same spirit of results in [LM05, FHHK24].

Claim 2.10. *Given an even integer $k \geqslant 2$ and functions $f : X \times Y \to [-1,1]$, $g : X \times Z \to [-1,1]$, $h : Y \times Z \to [-1,1]$, we have*

$$\Lambda(f,g,h) \leqslant \|f\|_{k/(k-1)}\|g\|_{U(2,k)}\|h\|_{U(2,k)}.$$

Proof. The proof follows from Hölder's inequality:

$$\begin{aligned}
\Lambda(f,g,h) &= \mathop{\mathbb{E}}_{x,y,z} f(x,y)g(x,z)h(y,z) \\
&= \mathop{\mathbb{E}}_{x,y} f(x,y)\left(\mathop{\mathbb{E}}_{z} g(x,z)h(y,z)\right) \\
&\leqslant \left(\mathop{\mathbb{E}}_{x,y} |f(x,y)|^{k/(k-1)}\right)^{(k-1)/k} \times \left(\mathop{\mathbb{E}}_{x,y}\left(\mathop{\mathbb{E}}_{z} g(x,z)h(y,z)\right)^k\right)^{1/k} \\
&\leqslant \|f\|_{k/(k-1)} \times \|g\|_{U(2,k)} \times \|h\|_{U(2,k)}
\end{aligned}$$

where the last inequality follows from Lemma 3.5. □

When f, g, h are indicators of sets of density δ, a common strategy is to decompose $g = \delta + g_0, h = \delta + h_0$, where $\mathbb{E}[g_0] = \mathbb{E}[h_0] = 0$. By linearity, there will be a main term $\Lambda(f, \delta, \delta) = \delta^3$ and various error terms. Suppose we try to bound $\Lambda(f, g_0, h_0)$ using Claim 2.10 with $k = 2$. We have

$$\Lambda(f, g_0, h_0) \leqslant \|f\|_2\|g_0\|_{U(2,2)}\|h_0\|_{U(2,2)} = \delta^{1/2}\|g_0\|_{U(2,2)}\|h_0\|_{U(2,2)}.$$

If this error terms exceeds $\Omega(\delta^3)$, then without loss of generality we can assume $\|g_0\|_{U(2,2)} \geqslant \Omega(\delta^{5/4})$. With some regularity conditions on the rows of g, this can be converted to a rectangle of density $\Omega(\delta^{10})$ where the density of g has increased from δ to $\delta + \Omega(\delta^{10})$ (see [Gre04, Proposition 5.7], for example). This density increment is quite weak; it requires $\Omega(1/\delta^{10})$ iterations before the density has increased by a constant factor, and so we can only guarantee the rectangle we are left with has density at least $\delta^{O(1/\delta^{10})}$.

On the other hand, if we set $k = \Omega(\log(1/\delta))$ sufficiently large, then we obtain

$$\Lambda(f, g_0, h_0) \leqslant \delta^{1-1/k}\|g_0\|_{U(2,k)}\|h_0\|_{U(2,k)} \leqslant 2\delta\|g_0\|_{U(2,k)}\|h_0\|_{U(2,k)}.$$

Now if $\Lambda(f, g_0, h_0) > \Omega(\delta^3)$, then without loss of generality we can assume $\|g_0\|_{U(2,k)} \geqslant \Omega(\delta)$. We will later show how to convert this[2] to $\|g\|_{U(2,k)} \geqslant (1 + \Omega(1))\delta$, at which point Lemma 2.9 gives a rectangle of density roughly δ^{2k+1} where the density of g has increased by a constant factor. This density increment process is more efficient, and it will eventually lead to quasi-polynomial bounds in δ.

[2] In actuality, we use $\Lambda(f, g_0, h_0) > \Omega(\delta^3)$ to deduce $\mathbb{E}_z[g(x,z)h(y,z)] - \delta^2$ is far from uniform, which in turn implies $\|g\|_{U(2,k)} \geqslant (1 + \Omega(\varepsilon))\delta$. The full details are present in Section 7, stated in the contrapositive. However, it is known how to perform the stated conversion (with some loss in parameters) under certain conditions (e.g. [FHHK24, Lemma 2.9]).

Despite this, applying Lemma 2.9 does not suffice to give strong enough bounds on the $(2,k)$-grid norms of the marginals of A. Suppose we applied Lemma 2.7 to obtain a cube C where the marginals of $A \cap C$ are $(O(d^2), \varepsilon)$-spread for some constant $\varepsilon > 0$. Since $|A| \geqslant 2^{-d}|S|$, this would give a cube C of size $|C| \geqslant 2^{-O(d^3)}N^3$ where $|A \cap C| \geqslant 2^{-(d+1)}|S \cap C|$. If we try to apply Lemma 2.9 with these parameters, we run into an issue: the marginals of A on faces of C can be very sparse. Since S is a weak permutation function which is pseudorandom against cubes, the density of A_{XZ} could scale with the density of C:

$$\mathbb{E}[A_{XZ}] = \frac{|A \cap C|}{|X||Z|} \leqslant \frac{|S \cap C|}{|X||Z|} \approx \frac{|S|}{N^3} \cdot \frac{|C|}{|X||Z|} \leqslant \frac{|Y|}{N}.$$

In particular, $|Y|$ could be as small as $2^{-\Omega(d^3)}N$, so at best we can only guarantee $\mathbb{E}[A_{XZ}] \geqslant 2^{-O(d^3)}$. Therefore, to bound

$$\|A_{XY}\|_{U(2,k)} \leqslant (1+2\varepsilon)\|A_{XY}\|_1,$$

we would need to guarantee that A_{XY} is $(\Omega(d^3 k), \varepsilon)$-spread. That is, to apply the theorem we require a far stronger assumption than the initial $(O(d^2), \varepsilon)$-spreadness. If we try to fix this by strengthening our spreadness assumption, the ambient cube becomes sparser, and we end up "chasing our own tail."

While the marginals of A to faces of C can be very sparse, they are dense inside of the marginals of S. Additionally, the marginals of S inherit strong pseudorandomness properties since S is pseudorandom against cubes. In particular, $A_{XY} \subset S_{XY}$, where S_{XY} is *pseudorandom against rectangles.*

Definition 2.11 (Pseudorandom against rectangles)**.** *A set $T \subset X \times Y$ is* γ-pseudorandom against rectangles *if for every rectangle $R \subset X \times Y$, we have*

$$\left| \mathop{\mathbb{E}}_{(x,y)\in T}[R(x,y)] - \mathop{\mathbb{E}}_{x\in X, y\in Y}[R(x,y)] \right| \leqslant \gamma.$$

Similar to pseudorandomness against cubes, pseudorandomness against rectangles guarantees that the density of T is roughly the same when restricting to any large rectangle. Now, if $S \subset [N]^3$ is pseudorandom against cubes, then the following claim shows that its marginals on any large cube are pseudorandom against rectangles, with a small loss in parameters. The proof is deferred to Section 3.

Claim 2.12. *Suppose $S \subset [N]^3$ is a weak permutation function which is γ-pseudorandom against cubes. Suppose $C = X \times Y \times Z$ is a cube of size $|C| \geqslant \gamma^{1/2}N^3$. Then, the marginals S_{XY}, S_{XZ}, S_{YZ} are $O(\gamma^{1/2})$-pseudorandom against rectangles.*

Our goal will be to exploit the fact that S_{XY} is pseudorandom against rectangles in order to argue that A_{XY} shares similar properties to sets which are dense in $X \times Y$. This situation is not uncommon; oftentimes, one can prove dense subsets of sparse pseudorandom sets satisfy similar properties to dense sets [KRSS10, CFZ14, CFZ15, CG16].

One possible strategy to overcome this obstacle is to apply Lemma 2.9 to some *globally* dense set $D \subseteq [N]^3$ that "models" A. In particular, one would want D's guaranteed density increment onto a large rectangle to imply a similar property for A, only with A's density being measured with respect to the pseudorandom set T. Such *dense model theorems* appear in various contexts throughout theoretical computer science, combinatorics, and number theory, perhaps most notably as a central ingredient in the proof of the celebrated Green-Tao theorem [GT08]. Unfortunately, the tradeoffs in standard formulations (see e.g. [RTTV08, Theorem 2.2]) are not quantitatively strong enough for our purposes, and we do not pursue this direction further. Instead, we proceed with a self-contained method of proof by directly modifying the steps in [KLM24, Lemma 4.7] to work in our pseudorandom setting. We are able to achieve a nearly identical lemma, but crucially the size of the provided rectangle depends on the function's density relative to a pseudorandom set rather than globally.

We need the following definitions (see Section 3 for the formal definitions). Let μ_T denote the uniform distribution over $T \subset X \times Y$. Given a non-negative function f supported on T, we have $\|f\|_{1(\mu_T)} = \mathbb{E}_{(x,y)\in T}[f(x,y)]$. The proof of the following lemma can be found in Section 6.

Lemma 2.13 (Relative version of [KLM24, Lemma 4.7])**.** *Let $f : X \times Y \to [0,1]$ be a function supported on a set $T \subset X \times Y$ of size $\tau|X||Y|$; suppose that $\|f\|_{1(\mu_T)} = \delta$. Let $\ell, k \in \mathbb{N}$. Additionally, assume that T is γ-pseudorandom with respect to rectangles for $\gamma \leqslant \varepsilon^8 \cdot (\tau\delta/2)^{O(\ell k)}$ small enough. If*

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

then there exists some rectangle $R \subset X \times Y$ with

$$\mathop{\mathbb{E}}_{(x,y)\in R} f(x,y) \geqslant \left(1 + \frac{\varepsilon}{64}\right)\|f\|_1 \quad \text{and} \quad \|R\|_1 \geqslant \frac{1}{64} \cdot \varepsilon \cdot \delta^{\ell k+1}.$$

It is essential to our argument that the density of the rectangle in the conclusion depends only on the density of f in T, rather than $\|f\|_1$.

2.3 A sparse von Neumann lemma

At this point, we have used Lemma 2.7 to obtain a cube $C = X \times Y \times Z$ where $A \cap C$ satisfies various pseudorandomness properties. Namely, the marginals of A on faces of C are spread. In the previous section, we saw that Lemma 2.13 will let us efficiently convert spreadness into bounded grid norms, even if the function is supported on a sparse pseudorandom set. The goal from here is to argue that the pseudorandom properties we obtained on C are enough to ensure that $A \cap C$ contains roughly the same number of patterns as a random set of the same density. Such results are often called *Generalized von Neumann lemmas.* In [KLM24], they prove the following:

Lemma 2.14 ([KLM24, Corollary 4.9])**.** *Let $f : X \times Y \to [0,1]$, $g : X \times Z \to [0,1], h : Y \times Z \to [0,1]$. Let $d \geqslant 1$ and $\varepsilon \in (0, 1/160)$, and set $r = \Omega((d^2 + d\log(1/\varepsilon))/\varepsilon)$. Assume that:*

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

Then

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

One can view this statement as being in the "dense" setting, where 2^{-d} is relatively large. Notice that for constant ε, the spreadness parameter r scales like $\Omega(d^2)$. We would like to obtain a similar conclusion with $f = A_{XY}, g = A_{XZ}, h = A_{YZ}$ which are all dense inside a pseudorandom set. If we apply Lemma 2.7 directly, we run into a similar issue as discussed in the previous section. Namely, the spreadness requirement scales with the density of f, g, h, and so we can never find a cube C where the marginals of A are both sufficiently spread and dense.

Thus, if we want to guarantee mixing for A_{XY}, A_{XZ}, A_{YZ}, we need a version of Lemma 2.14 where the spreadness requirement scales with the density of $A \cap C$ in $S \cap C$, rather than the density of the ambient cube C. If we look a bit into the proof of Lemma 2.14, we find where the issue lies. Define

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

A key step in their proof is to apply Hölder's inequality, then try to control

$$\|g \circ h - \mathbb{E}[g]\,\mathbb{E}[h]\|_p \leqslant (1 + O(1))\,\mathbb{E}[g]\,\mathbb{E}[h]$$

where p depends on the $\mathbb{E}[h]$. They do this by controlling the grid norms

$$\|g\|_{U(2,k)} \leqslant (1 + O(1))\,\mathbb{E}[g] \quad \text{and} \quad \|h\|_{U(2,k)} \leqslant (1 + O(1))\,\mathbb{E}[h].$$

for $k = O(\log(1/\mathbb{E}[f]))$. As we saw before, $f = A_{XY}$ could be very sparse, and so controlling $(2,k)$-grid norms would require spreadness which is not feasible in our setting.

The choice of k in the above approach is far too large for our purposes. To remedy this, we will use the fact that S is pseudorandom against cubes. It follows from Claim 2.12 that the marginals of S to faces of C are pseudorandom against rectangles with a small loss in parameters. The main observation is that for any fixed $z \in Z$, the function $g(x,z)h(y,z)$ is a soft rectangle. In particular, Claim 3.2 implies $g \circ h = \mathbb{E}_z\, g(x,z)h(y,z)$, and therefore $(g \circ h)^k$, is a convex combination of rectangles. The pseudorandomness of S_{XY} then gives that the k-norm of $g \circ h$ as a function on $X \times Y$ is within a small additive error of the k-norm of $g \circ h$ restricted to S_{XY}. Thus, it suffices to bound the k-norm of $g \circ h$ *viewed as a function restricted to* S_{XY}. Inside of S_{XY}, the function f will be considerably denser. We can then choose k to depend solely on the density of f in S_{XY}, rather than the global density of f in $X \times Y$. The details of this argument and proof of the following lemma are given in Section 7.

Lemma 2.15 (Sparse von Neumann)**.** *Let $T \subset X \times Y$ be a set which is γ-pseudorandom against rectangles, and let $A \subset T$ be a set of size $|A| \geqslant 2^{-d}|T|$. Let $g : X \times Z \to [0,1], h : Y \times Z \to [0,1]$ be functions. Let $d \geqslant 1, \varepsilon \in (0, 1/20)$. For $k = O(d/\varepsilon)$ a large enough integer, suppose that*

1. $\|g\|_{U(2,k)} \leqslant (1+\varepsilon)\|g\|_1$,
2. $\|h\|_{U(2,k)} \leqslant (1+\varepsilon)\|h\|_1$,
3. *g, h are ε-left lower-bounded,*
4. *$\gamma \leqslant (\varepsilon \|g\|_1 \|h\|_1)^{O(d)}$ is small enough.*

Then

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

2.4 Putting everything together

We now have the tools to prove our main theorem.

Proof of Theorem 2.4. The proof consists of three main steps. First, we will apply Lemma 2.7 to restrict A to some large cube $C = X \times Y \times Z$ so that the marginals A_{XY}, A_{XZ}, A_{YZ} are sufficiently spread. We will then use Lemma 2.13 to show that spreadness is sufficient to imply the marginals of A are uniform in an appropriate grid norm. Finally, we will lower bound the number of patterns $(x,y) \in A_{XY}, (x,z) \in A_{XZ}, (y,z) \in A_{YZ}$ by applying Lemma 2.15.

Obtaining spreadness. Let $\varepsilon > 0$ be a small enough absolute constant, to be determined later. We first apply Lemma 2.7 with $r = c_1 d^2/\varepsilon$ for some large enough constant $c_1 > 0$, which we can do since $\gamma \leqslant 2^{-O(d^3)}$ is sufficiently small. We find there exists some cube $C = X \times Y \times Z$ of density at least $2^{-O(d^3)}$ with

1. $|A \cap C| \geqslant 2^{-(d+1)}|S \cap C|$.
2. The marginals A_{XZ}, A_{YZ} are $(c_1 d^2/\varepsilon, \varepsilon/64)$-spread.
3. The marginals A_{XZ}, A_{YZ} are $(\varepsilon/64)$-left lower-bounded.

Obtaining uniformity from spreadness. We will bound the grid norm for A_{XZ}; the proof for A_{YZ} is similar. Assume for the sake of contradiction that

$$\|A_{XZ}\|_{U(2,k)} > (1+\varepsilon)\,\mathbb{E}[A_{XZ}]$$

for $k = O(d/\varepsilon)$ a large enough integer. By Claim 2.12, we can infer that S_{XZ} is $(c_2\gamma^{1/2})$-pseudorandom against rectangles for some constant $c_2 > 0$. For $\gamma \leqslant 2^{-O(d^4)}$ sufficiently small, we have $c_2\gamma^{1/2} \leqslant \varepsilon^8 \cdot 2^{-O(d^3 k)}$, so we can apply Lemma 2.13 to obtain a rectangle R with

$$\mathop{\mathbb{E}}_{(x,z)\in R}[A_{XZ}] > \left(1+\frac{\varepsilon}{64}\right)\mathbb{E}[A_{XZ}] \quad \text{and} \quad \frac{|R|}{|X||Z|} \geqslant \frac{\varepsilon}{64}\cdot 2^{-O(dk)}.$$

Since $k = O(d/\varepsilon)$, this contradicts our assumption that A_{XZ} is $\left(c_1 d^2/\varepsilon, \varepsilon/64\right)$-spread for $c_1 > 0$ large enough.

Counting triples. At this point, we have established the following conditions:

1. $\|A_{XZ}\|_{U(2,k)} \leqslant (1+\varepsilon)\cdot\mathbb{E}[A_{XZ}]$,
2. $\|A_{YZ}\|_{U(2,k)} \leqslant (1+\varepsilon)\cdot\mathbb{E}[A_{YZ}]$,
3. A_{XZ}, A_{YZ} are $(\varepsilon/64)$-left lower-bounded.

In order to apply Lemma 2.15, it remains to check that S_{XY} is sufficiently pseudorandom against rectangles in terms of the density of A_{XZ}, A_{YZ}. By Claim 2.12, we know that S_{XY} is $(c_2\gamma^{1/2})$-pseudorandom against rectangles for some constant $c_2 > 0$. We already established that $|A \cap C| \geqslant 2^{-d}|S \cap C|$. Since S is γ-pseudorandom against cubes, we have

$$|S \cap C| \geqslant \frac{|S||C|}{N^3} - \gamma|S| \geqslant \frac{2^{-s}}{N}|C| - \gamma N^2 \geqslant 2^{-O(d^3+s)}N^2,$$

where the inequality holds for $\gamma \leqslant 2^{-O(d^3+s)}$ small enough. Thus, we have

$$\mathbb{E}[A_{XY}] = \frac{|A\cap C|}{|X||Y|} \geqslant \frac{2^{-d}|S\cap C|}{|X||Y|} \geqslant 2^{-O(d^3+s)}.$$

The same inequality holds for $\mathbb{E}[A_{XZ}], \mathbb{E}[A_{YZ}]$. For $\gamma \leqslant 2^{-O(d^4+ds)}$ sufficiently small, we have $c_2\gamma^{1/2} \leqslant \varepsilon^{O(d)}\cdot 2^{-O(d^4+ds)} \leqslant (\varepsilon\cdot\mathbb{E}[A_{XZ}]\,\mathbb{E}[A_{YZ}])^{O(d)}$. The conditions for Lemma 2.15 are satisfied, which gives

$$\mathop{\mathbb{E}}_{x\in X, y\in Y, z\in Z}[A_{XY}(x,y)A_{XZ}(x,z)A_{YZ}(y,z)] = (1\pm O(\varepsilon))\,\mathbb{E}[A_{XY}]\,\mathbb{E}[A_{XZ}]\,\mathbb{E}[A_{YZ}].$$

We now choose ε so that the implicit $O(\varepsilon)$ term appearing on the right hand side is at most 1/2. Since $|C| \geqslant 2^{-O(d^3)}N^3$, this gives at least

$$(1-O(\varepsilon))\,2^{-O(d^3+s)}|C| \geqslant 2^{-O(d^3+s)}N^3$$

many points $(x,y,z)\in C$ with the property that $(x,y)\in A_{XY}, (x,z)\in A_{XZ}, (y,z)\in A_{YZ}$. □

3 Preliminaries

Given positive numbers x and y, we shorthand $x = y\pm\varepsilon$ for $y-\varepsilon \leqslant x \leqslant y+\varepsilon$.

Sets. Let X, Y, Z, and occasionally $\mathcal{X}$ be finite sets throughout. We define the natural numbers to exclude zero; that is, $\mathbb{N} = \{1, 2, \ldots\}$.

Asymptotics. We use standard asymptotic notation of $O(\cdot)$, $\Omega(\cdot)$, and $\Theta(\cdot)$ to suppress fixed constants that do not depend on any parameters.

Distributions. A distribution ν over $\mathcal{X}$ is a non-negative function $\nu : \mathcal{X} \to \mathbb{R}_{\geqslant 0}$ with $\mathbb{E}[\nu] = 1$. For $f : \mathcal{X} \to \mathbb{R}_{\geqslant 0}$ define $\nu(f) = \mathbb{E}_{x\in\mathcal{X}}\,\nu(x)f(x)$ to be the average of f under ν. Additionally, let μ_X be the uniform distribution over X.

Functions. For functions $f, g : \mathcal{X} \to \mathbb{R}$, we define inner products and norms with the normalized counting measure on $\mathcal{X}$, namely

$$\langle f, g \rangle = \mathop{\mathbb{E}}_{x \in \mathcal{X}} f(x)g(x) \quad \text{and} \quad \|f\|_p = \left(\mathop{\mathbb{E}}_{x \in \mathcal{X}} |f(x)|^p \right)^{1/p} \quad \text{for } 1 \leqslant p < \infty,$$

as well as $\|f\|_\infty = \max_{x \in \mathcal{X}} |f(x)|$. When $\|f\|_\infty \leqslant 1$, we will refer to f as being 1-bounded. We will also want to work with other distributions on $\mathcal{X}$. For a distribution ν on $\mathcal{X}$, we write

$$\langle f, g \rangle_\nu = \mathop{\mathbb{E}}_{x \in \mathcal{X}} \nu(x) f(x) g(x) \quad \text{and} \quad \|f\|_{p(\nu)} = \left(\mathop{\mathbb{E}}_{x \in \mathcal{X}} \nu(x) |f(x)|^p \right)^{1/p} \quad \text{for } 1 \leqslant p < \infty.$$

For convenience, we will often overload notation when working with a set S by letting $S(x)$ denote its indicator function $\mathbb{1}(x \in S)$.

3.1 Permutation functions

Definition 2.2 ((Weak) permutation function)**.** *Let $f : [N]^3 \to \{0, 1\}$. We call f a* permutation function *if for every fixing of any two coordinates, there is exactly one fixing of the remaining coordinate so that $f(x, y, z) = 1$. If instead there exists at most one value, we call f a* weak permutation function*.*

Occasionally, we refer to a product set as a (weak) permutation function if the corresponding indicator functions satisfy the definition. Note that these definitions coincide with those of *2-dimensional permutations* and *linjections*, respectively, from [LL14, LPS19] (both of which generalize *(weak) graph functions* [BDPW10, Shr18]).

We often work with sets of a particular form, where the following notation will be convenient.

Notation 2.3. *For a set $A \subset [N]^3$ and a cube $C = X \times Y \times Z$, let $A_{XY} \subset X \times Y$ denote the projection of A onto its XY-face. More formally, a point $(x, y) \in A_{XY}$ if and only if there exists $z \in Z$ with $(x, y, z) \in A \cap C$. The sets A_{XZ} and A_{YZ} are defined analogously.*

Given a marginal such as A_{XY}, it will always be clear from context what the ambient cube $C = X \times Y \times Z$ is. When A is a (weak) permutation function (as will typically be the case for us), the projections satisfy $|A \cap C| = |A_{XY}| = |A_{XZ}| = |A_{YZ}|$.

3.2 Rectangles and pseudorandomness

Definition 3.1 ((Soft) rectangle)**.** *A* rectangle *is a function of the form $f(x)g(y)$, where $f : X \to \{0, 1\}$ and $g : Y \to \{0, 1\}$. If we relax the codomains of f and g to be $[0, 1]$, we call $f(x)g(y)$ a* soft rectangle*.*

We will often call a product set $X \times Y$ a rectangle, viewing it as the product of indicator functions $X(x)Y(y)$. Additionally, we extend these notions in three dimensions to *cubes*. Soft rectangles have the following convenient property.

Claim 3.2 ([KLM24, Claim 4.5])**.** *Let $f : X \to [0, 1]$ and $g : Y \to [0, 1]$. The soft rectangle $f(x)g(y)$ can be written as a convex combination of rectangles.*

A key pseudorandomness notion in this work is *pseudorandomness against cubes.*

Definition 2.1 (Pseudorandom against cubes)**.** *A set $S \subset X \times Y \times Z$ is* γ-pseudorandom against cubes *if for every cube $C \subset X \times Y \times Z$, we have*

$$\left| \mathop{\mathbb{E}}_{(x,y,z) \in S} [C(x, y, z)] - \mathop{\mathbb{E}}_{x \in X, y \in Y, z \in Z} [C(x, y, z)] \right| \leqslant \gamma.$$

A similar notion appeared in [KLM24] with a multiplicative error term and the condition that C is a large enough cube. They also added a condition on the marginals of S, which becomes useful when S is a dense function. In this work, S is always a (weak) permutation function, so this was extra condition was not necessary.

We can also define *pseudorandomness against rectanngles* in a similar manner.

Definition 2.11 (Pseudorandom against rectangles)**.** *A set $T \subset X \times Y$ is* γ-pseudorandom against rectangles *if for every rectangle $R \subset X \times Y$, we have*

$$\left| \mathop{\mathbb{E}}_{(x,y)\in T}[R(x,y)] - \mathop{\mathbb{E}}_{x\in X, y\in Y}[R(x,y)] \right| \leqslant \gamma.$$

If $S \subset [N]^3$ is pseudorandom against cubes, then we can show that the marginals to faces of a large cube will be pseudorandom against rectangles with a small loss in parameters.

Claim 2.12. *Suppose $S \subset [N]^3$ is a weak permutation function which is γ-pseudorandom against cubes. Suppose $C = X \times Y \times Z$ is a cube of size $|C| \geqslant \gamma^{1/2}N^3$. Then, the marginals S_{XY}, S_{XZ}, S_{YZ} are $O(\gamma^{1/2})$-pseudorandom against rectangles.*

Proof. We will prove the statement for $T := S_{XY} \subset X \times Y$, the other marginals follow similarly. Suppose $C = X \times Y \times Z$, and let $R = X' \times Y' \subset X \times Y$ be some rectangle. We want to show that

$$\left| \mathop{\mathbb{E}}_{(x,y)\in T}[R(x,y)] - \mathop{\mathbb{E}}_{x\in X, y\in Y}[R(x,y)] \right| \leqslant O(\gamma^{1/2}).$$

Let $C' = X' \times Y' \times Z$. Since S is a weak permutation function, this is equivalent to showing

$$\left| \frac{|S \cap C'|}{|S \cap C|} - \frac{|C'|}{|C|} \right| \leqslant O(\gamma^{1/2}).$$

By the γ-pseudorandomness of S, we have

$$\frac{|S \cap C|}{|S|} = \frac{|C|}{N^3} \pm \gamma \quad \text{and} \quad \frac{|S \cap C'|}{|S|} = \frac{|C'|}{N^3} \pm \gamma.$$

This gives

$$\frac{|C'| - \gamma N^3}{|C| + \gamma N^3} \leqslant \frac{|S \cap C'|}{|S \cap C|} \leqslant \frac{|C'| + \gamma N^3}{|C| - \gamma N^3}.$$

For $|C| \geqslant \gamma^{1/2}N^3$, we have

$$\frac{|S \cap C'|}{|S \cap C|} = \frac{|C'|}{|C|} \pm O(\gamma^{1/2}).$$

□

We will also want two other notions of pseudorandomness for functions on a rectangle, *spreadness* and *left lower-boundedness*. Both were introduced in [KLM24].

Definition 2.5 (Spread)**.** *Let $r \geqslant 1$ and $\varepsilon \in (0,1)$. A function $f : X \times Y \to [0,1]$ is* (r,ε)-spread *if for any rectangle $R = X' \times Y' \subset X \times Y$ of size $|R| \geqslant 2^{-r}|X||Y|$, we have*

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

In short, spreadness guarantees that the density of a given function cannot increase by restricting to some large rectangle. Along similar lines, left lower-boundedness will guarantee that the rows of f are not too sparse.

Definition 2.6 (Left lower-bounded)**.** *Let $\varepsilon \in (0,1)$. A function $f : X \times Y \to [0,1]$ is* ε-left lower-bounded *if for every $x \in X$, we have*

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

3.3 Grid norms

The *grid norm* is an analytic quantity that captures captures rectangular structure.

Definition 2.8 (Grid norms)**.** *For a function $f : X \times Y \to \mathbb{R}$ and $\ell, k \in \mathbb{N}$, let*

$$\begin{aligned} U_{\ell,k}(f) &= \mathop{\mathbb{E}}_{x_1,\ldots,x_\ell \in X} \left(\mathop{\mathbb{E}}_{y \in Y} f(x_1, y) \cdots f(x_\ell, y) \right)^k \\ &= \mathop{\mathbb{E}}_{y_1,\ldots,y_k \in Y} \left(\mathop{\mathbb{E}}_{x \in X} f(x, y_1) \cdots f(x, y_k) \right)^\ell \\ &= \mathop{\mathbb{E}}_{\substack{x \in X^\ell \\ y \in Y^k}} \prod_{i=1}^{\ell} \prod_{j=1}^{k} f(x_i, y_j). \end{aligned}$$

*The (ℓ, k)-*grid norm *of f is given by $\|f\|_{U(\ell,k)} := |U_{\ell,k}(f)|^{1/\ell k}$.*

The reader will notice that the $(2, 2)$-grid norm is exactly the box-norm. These norms were introduced in [KLM24] as a generalization of the classic box-norm from combinatorics. There it was shown that functions with large grid norms admit density increments when restricting to some large rectangle.

Lemma 2.9 ([KLM24, Lemma 4.7])**.** *Let $f : X \times Y \to [0, 1]$; suppose that $\|f\|_1 \geqslant \delta$. Let $\ell, k \in \mathbb{N}$. If*

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

then there exists some rectangle $R \subset X \times Y$ with

$$\mathop{\mathbb{E}}_{(x,y) \in R} f(x, y) \geqslant \left(1 + \frac{\varepsilon}{2}\right) \|f\|_1 \quad \textit{and} \quad \|R\|_1 \geqslant \frac{1}{2} \cdot \varepsilon \cdot \delta^{\ell k+1}.$$

The reader will notice that the contrapositive of Lemma 2.9 states that if f is spread (see Definition 2.5), then f has a bounded grid-norm. Along these lines, the following lemma will also let us convert pseudorandomness against rectangles into bounded grid norms.

Lemma 3.3. *Let $T \subset X \times Y$ be a set with size $|T| = \tau |X||Y|$ which is γ-pseudorandom against rectangles, and let $\ell, k \in \mathbb{N}$. If $\gamma < \frac{1}{4} \cdot \varepsilon^2 \cdot \tau^{\ell k+1}$, then $\|T\|_{U(\ell,k)} \leqslant (1 + \varepsilon)\tau$.*

Proof. Assume for the sake of contradiction that $\|T\|_{U(\ell,k)} \geqslant (1 + \varepsilon)\tau$. By Lemma 2.9, there exists a rectangle R with

$$\frac{|R \cap T|}{|R|} \geqslant \left(1 + \frac{\varepsilon}{2}\right) \frac{|T|}{|X \times Y|} \quad \text{and} \quad \frac{|R|}{|X \times Y|} \geqslant \frac{1}{2} \cdot \varepsilon \cdot \tau^{\ell k+1}.$$

This contradicts our assumption on γ, as pseudorandomness against rectangles of T implies

$$\gamma \geqslant \frac{|R \cap T|}{|T|} - \frac{|R|}{|X \times Y|} \geqslant \left(1 + \frac{\varepsilon}{2}\right) \frac{|R|}{|X \times Y|} - \frac{|R|}{|X \times Y|} \geqslant \frac{1}{4} \cdot \varepsilon^2 \cdot \tau^{\ell k+1}. \qquad \square$$

Before proceeding, we collect a number of useful facts about grid norms. Although technically, $\|\cdot\|_{U(\ell,k)}$ is not a norm, it is a semi-norm in the case where ℓ and k are both even [Hat10, Theorems 2.8, 2.9]. (Thus, we may apply a triangle inequality in that setting.) Much like standard k-norms, grid norms are monotonic.

Claim 3.4 ([KLM24, Claim 4.2])**.** *Let $\ell, \ell', k, k' \in \mathbb{N}$, where $\ell \leqslant \ell'$ and $k \leqslant k'$. Additionally, let $f : X \times Y \to \mathbb{R}_{\geqslant 0}$. Then,*

$$\|f\|_{U(\ell,k)} \leqslant \|f\|_{U(\ell',k)} \quad \textit{and} \quad \|f\|_{U(\ell,k)} \leqslant \|f\|_{U(\ell,k')}.$$

They may also be used to decouple two functions via an application of the Cauchy-Schwarz inequality.

Lemma 3.5 ([KLM24, Lemma 4.3])**.** *Let $g : X \times Z \to \mathbb{R}$ and $h : Y \times Z \to \mathbb{R}$. For even $k \in \mathbb{N}$ we have*

$$\mathop{\mathbb{E}}_{x \in X, y \in Y} \left(\mathop{\mathbb{E}}_{z \in Z} g(x,z)h(y,z) \right)^k \leqslant U_{2,k}(g)^{1/2} \cdot U_{2,k}(h)^{1/2}.$$

4 Applications

In this section, we present several applications of Theorem 2.4 to communication complexity and extremal combinatorics. The applications are chosen to illustrate the flavor of results one may obtain via our techniques, but the list is not exhaustive. Each subsection is self-contained and may be skipped according to the reader's preferences.

4.1 Communication complexity of permutation functions

Our first application is to communication complexity. Before providing the details, we briefly review the necessary setup. The *communication complexity* of a function f is the fewest number of bits required for a protocol to evaluate f. We will be exclusively interested in the *three-player number-on-forehead* (3-NOF) model of communication, where each player's input is viewed as being on their forehead, so that they may see all inputs except their own.

Definition 4.1 (Cylinder intersection)**.** *A set $A \subset [N]^3$ is a* cylinder intersection *if*

$$A = \left\{ (x,y,z) \in [N]^3 : (x,y) \in S_1, (x,z) \in S_2, (y,z) \in S_3 \right\}$$

for some sets $S_1, S_2, S_3 \subset [N]^2$.

Notice that the left-hand side of Theorem 2.4's conclusion is the size of a specific cylinder intersection. The work of [KLM24] referred to this as the "cylinder intersection closure of A," i.e. the smallest cylinder intersection containing A. In particular, when A is a cylinder intersection, the cylinder intersection closure of A is A itself, so this gives a lower bound on the size of cylinder intersections which can be contained in a pseudorandom set. It is well known that one can translate complexity information to combinatorial information in the form of cylinder intersections [BNS89].

Fact 4.2. *If there exists a b-bit non-deterministic protocol to determine membership in a set S, then S can be written as a union of 2^b cylinder intersections.*

For additional background, see, for example, the recent excellent book [RY20] on communication complexity.

Communication bounds on weak permutation functions are known to have strong relationships with bounds on a number of landmark combinatorial problems, such as corners [CFL83, LS21b], combinatorial lines (i.e. Hales-Jewett theorems) [CFL83, Shr18], and dense Ruzsa-Szemerédi graphs [LPS19, AS20]. These functions (and several variants) have been explicitly considered in a number of prior works [BDPW10, LL14, LPS19, AS20]. Most relevant to our results, Linial, Pitassi, and Shraibman proved the deterministic communication complexity for permutation functions is $\Omega(\log \log \log N)$ [LPS19]. (One may wish to contrast this with the existence of a simple constant communication randomized protocol via reduction to equality.) The following theorem shows that this bound can be substantially improved in the case of sufficiently pseudorandom functions. We state the theorem for permutation functions, but it can be easily adapted to work for (dense enough) weak permutation functions.

Theorem 4.3. *Suppose $S \subset [N]^3$ is a permutation function which is N^{-c}-pseudorandom against cubes. Then, any non-deterministic 3-NOF protocol for determining membership in S requires at least $\Omega((c \log N)^{1/4})$ bits of communication.*

Proof. Suppose there exists a b-bit non-deterministic protocol to determine membership in S. By Fact 4.2, S can be written as a union of 2^b cylinder intersections $A_1, \ldots, A_{2^b} \subset S$. Let A be the largest one, where $|A| \geqslant 2^{-b}N^2$. Assume towards a contradiction that $b = O((c \log N)^{1/4})$ small enough. Then $N^{-c} \leqslant 2^{O(b^4)}$ and we may apply Theorem 2.4 (with $d = b, s = 1$) and deduce that

$$A' = \{(x, y, z) \in [N]^3 : (x, y) \in A_{XY}, (x, z) \in A_{XZ}, (y, z) \in A_{YZ}\}$$

has size $|A'| \geqslant 2^{-O(b^3)}N^3$. However, since A is a cylinder intersection, we have $A' = A$. We thus reached a contradiction since $|A| \leqslant |S| = N^2$. □

As a corollary, we get a lower bound for Exactly-N in quasirandom groups such as $G = \mathrm{SL}_2(\mathbb{F}_p)$, discussed in Section 4.2.

Theorem 1.1 (Special case of Theorem 4.3)**.** *Any non-deterministic 3-NOF protocol computing* Exactly-N *over* $G = \mathrm{SL}_2(\mathbb{F}_p)$ *for prime* p *requires* $\Omega(\log^{1/4}|G|)$ *bits of communication.*

Proof. Let $G = \mathrm{SL}_2(\mathbb{F}_p)$. The proof follows from Theorem 4.3 and the fact that $S = \{(x, y, z) \in G : xyz = 1_G\}$ is $D^{-1/2}$-pseudorandom against cubes for $D = (p-1)/2$ (see Lemma 4.8). □

4.2 Corners in quasirandom groups

The notion of quasirandom groups was introduced by Gowers [Gow08] in studying product-free sets. Quasirandom groups enjoy the property that for any two large sets $A, B \subset G$, the distribution obtained by taking uniform random samples $a \in A$, $b \in B$ and outputting ab is close to uniform in an L_2-sense. Of course, abelian groups fail to satisfy the mixing property mentioned above. In $\mathbb{F}_2^n$ for instance, a subspace V of codimension 1 is very dense, but $V + V = V$ is far from being uniform. We now present the formal definition.

Definition 4.4 (D-Quasirandom group)**.** *A finite group G is D-quasirandom if every nontrivial irreducible representation over $\mathbb{C}$ has dimension at least D.*

One can show that every irreducible representation of a finite group G must have dimension at most $\sqrt{|G|}$. We record this fact for later use, which can be found in most introductory texts on representation theory of finite groups. For example, this follows from Proposition 5 in [Ser77, Chapter 2.4].

Fact 4.5. *Every complex irreducible representation of a finite group G has dimension at most $\sqrt{|G|}$.*

There are choices of groups G which are D-quasirandom for $D = |G|^{\Omega(1)}$. Contrast this with abelian groups, where every irreducible representation has dimension 1. A common example of a quasirandom group is

$$\mathrm{SL}_2(\mathbb{F}_p) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a, b, c, d \in \mathbb{F}_p, ad - bc = 1 \right\},$$

the set of 2×2 matrices over the finite field $\mathbb{F}_p$ for p prime with determinant 1. A classical result of Frobenius [Fro68] shows that every nontrivial irreducible representation of $\mathrm{SL}_2(\mathbb{F}_p)$ has dimension at least $\frac{p-1}{2} \sim |G|^{1/3}$. (For exposition of these facts and additional background on quasirandom groups, see e.g. [Tao11].)

Quasirandom groups have seen various application in constructing pseudorandom objects. The Ramanujan graphs of Lubotzky, Phillips, and Sarnak [LPS88] are built from Cayley graphs of $\mathrm{PSL}_2(\mathbb{F}_p)$ for specific choices of p. (The group $\mathrm{PSL}_2(\mathbb{F}_p)$ is obtained by quotienting $\mathrm{SL}_2(\mathbb{F}_p)$ by its center.) Bourgain and Gamburd [BG08] showed that random Cayley graphs on $\mathrm{SL}_2(\mathbb{F}_p)$ are expanders. The quasirandomness property was useful for arguing about the multiplicity of eigenvalues of the Cayley graph. Outside of graph theory, ideas from the study of quasirandom groups led to an optimal inapproximability result for k-LIN over non-abelian groups [BK21]. A recent work of Derksen, Lee, and Viola [DLV24] building on [GV19] proved k-NOF lower bounds for computing an "interleaved product" of elements coming from a quasirandom group. In particular, their bounds are best understood in the regime where k is growing, and they match that of [BNS89].

Before we can formally state the mixing property of quasirandom groups, we define the convolution of two functions $f, g\colon G \to \mathbb{R}$ as

$$(f * g)(x) = \mathop{\mathbb{E}}_{y\in G}\left[f(y)g(y^{-1}x)\right].$$

Notice that the arguments to the functions in the expectation $y, y^{-1}x$ satisfy $y \cdot (y^{-1}x) = x$. When μ_1, μ_2 are distributions, then $(\mu_1 * \mu_2)(x)$ is the probability of independent samples $z_1 \sim \mu_1$ and $z_2 \sim \mu_2$ satisfying $x = z_1 z_2$. The most useful property of quasirandom groups is that convolutions of large sets mix. More formally, we have the following:

Theorem 4.6 ([Gow08, BNP08]). *Let G be a D-quasirandom group and $f, g : G \to \mathbb{R}$, and assume at least one has mean zero. Then*

$$\|f * g\|_2 \leqslant D^{-1/2}\|f\|_2\|g\|_2.$$

There are various proofs of this fact [BNP08, Gow08, Tao11] which use representation theory. Over abelian groups, convolutions and L_2-norms have pleasant interpretations when working in the Fourier basis. Many of these statements have clean analogs over non-abelian groups which can be formulated using representation theory.

Theorem 4.6 can be used to count solutions to equations in quasirandom groups. [Gow08] did exactly this to argue about the size of product-free sets in quasirandom groups. Namely, we can count solutions (x, y) where $x \in X, y \in Y, xy \in Z$. We have the following:

Corollary 4.7 ([Gow08, BNP08]). *Let G be a D-quasirandom group, and suppose $X, Y, Z \subset G$ are subsets. Then,*

$$\left|\mathop{\mathbb{E}}_{x,y\in G}[X(x)Y(y)Z(xy)] - \mathbb{E}[X]\,\mathbb{E}[Y]\,\mathbb{E}[Z]\right| \leqslant D^{-1/2}\|X\|_2\|Y\|_2\|Z\|_2.$$

Proof. Let $f = X - \mathbb{E}[X]$, $g = Y - \mathbb{E}[Y]$, and $h = Z - \mathbb{E}[Z]$. We have

$$\begin{aligned}
\mathop{\mathbb{E}}_{x,y\in G}[X(x)Y(y)Z(xy)] &= \mathop{\mathbb{E}}_{z\in G}[(X * Y)(z)Z(z)] \\
&= \mathbb{E}[X]\,\mathbb{E}[Y]\,\mathbb{E}[Z] + \mathop{\mathbb{E}}_{z\in G}[(f * g)(z)h(z)] \\
&= \mathbb{E}[X]\,\mathbb{E}[Y]\,\mathbb{E}[Z] \pm \|f * g\|_2\|h\|_2 && \text{(Cauchy-Schwarz)} \\
&= \mathbb{E}[X]\,\mathbb{E}[Y]\,\mathbb{E}[Z] \pm D^{-1/2}\|f\|_2\|g\|_2\|h\|_2. && \text{(Theorem 4.6)}
\end{aligned}$$

Since f and $\mathbb{E}[X]$ are orthogonal, we have $\|f\|_2 = \left(\mathbb{E}[X^2] - \mathbb{E}[X]^2\right)^{1/2} \leqslant \|X\|_2$, and similarly for g and h. $\square$

We can essentially rephrase Corollary 4.7 as a statement about pseudorandomness against cubes.

Lemma 4.8. *Let G be a D-quasirandom group. Then the set*

$$S \coloneqq \{(x, y, xy) : x, y \in G\}$$

is $D^{-1/2}$-pseudorandom against cubes.

Proof. If $C = X \times Y \times Z \subset G^3$ is a cube, then Corollary 4.7 implies

$$\left|\mathop{\mathbb{E}}_{(x,y,z)\in S}[C(x,y,z)] - \mathop{\mathbb{E}}_{x,y,z\in G}[C(x,y,z)]\right| = \left|\mathop{\mathbb{E}}_{x,y\in G}[X(x)Y(y)Z(xy)] - \mathbb{E}[X]\,\mathbb{E}[Y]\,\mathbb{E}[Z]\right| \leqslant D^{-1/2}. \quad \square$$

Note that the same proof also works to show that $\{(x, y, z) : xyz = 1_G\}$ is pseudorandom against cubes by replacing Z in the above proof with $Z^{-1} = \{z^{-1} : z \in Z\}$. Together with Lemma 4.8, we obtain our result on corner-free sets in quasirandom groups as a corollary of Theorem 2.4, where we define corners in the following way:

Definition 4.9 (Corner)**.** *Let G be a finite group. A* corner *is a triple $\{(x,y),(xg,y),(x,gy)\} \subset G \times G$. A corner is* nontrivial *if $g \neq 1_G$.*

Corollary 4.10. *Let G be a D-quasirandom group. Then, any corner-free subset of $G \times G$ has size at most*

$$\exp\left(-\Omega\left(\log^{1/4} D\right)\right) \cdot |G|^2.$$

Proof. Let $B \subset G \times G$ be a corner-free set of size $|B| \geqslant 2^{-d}|G|^2$ for some $d \geqslant 1$. Assume for the sake of contradiction that $d \leqslant O(\log^{1/4} D)$ for some small enough implicit constant. Define $S \subset G \times G \times G$ to be

$$S := \{(x,y,xy) : x,y \in G\}.$$

Notice that by the group property, S is a permutation function. Additionally, Lemma 4.8 implies that S is $D^{-1/2}$-pseudorandom with respect to cubes. We will embed B as a subset of S by defining

$$A := \{(x,y,xy) : (x,y) \in B\} \subset S.$$

Observe A has size $|A| \geqslant 2^{-d}|S|$. Additionally,

$$\{(x,y),(x,x^{-1}z),(zy^{-1},y)\} \subset B \iff \{(x,y,xy),(x,x^{-1}z,z),(zy^{-1},y,z)\} \subset A.$$

(Note that $\{(x,y),(x,x^{-1}z),(zy^{-1},y)\}$ corresponds to a corner by an appropriate change of variables.) This along with the fact that A is a weak permutation function implies that the corner count in B is given by

$$\sum_{x,y,z \in G} A_{XY}(x,y)A_{XZ}(x,z)A_{YZ}(y,z),$$

which since B has no nontrivial corners, is at most the number of trivial corners $|G|^2$. We can use Theorem 2.4 to lower bound this count as long as we can verify that S is sufficiently pseudorandom against cubes. Indeed, for $d \leqslant O(\log^{1/4} D)$ small enough, we have $D^{-1/2} \leqslant 2^{-O(d^4)}$. By Fact 4.5, we have $D \leqslant |G|^{1/2}$. This gives a lower bound on the corner count of

$$2^{-O(d^3)}|G|^3 \geqslant 2^{-O(\log^{3/4} D)}|G|^3 > |G|^2$$

which gives the desired contradiction. ☐

4.3 Improved triangle removal lemma for pseudorandom graphs

An important result in extremal combinatorics is the triangle removal lemma [RS78], which says that any N-vertex graph which is ε-far from being triangle-free[3] contains at least δN^3 triangles, where $\delta = \delta(\varepsilon)$ does not depend on N. The original proof of Ruzsa and Szemerédi is based on Szemerédi's regularity lemma [Sze75], which gives very poor quantitative bounds on δ as a function of ε. Despite much effort towards obtaining improved bounds, the best known bound [Fox11] is still quantitatively very weak - δ^{-1} is at most a tower of exponentials of height about $\log(1/\varepsilon)$. To contrast that, the best known lower bound on δ is quasi-polynomial [Alo02], and shows that $\delta \leqslant \varepsilon^{O(\log(1/\varepsilon))}$ is necessary.

Our main result can be seen as a quasi-polynomial upper bound for the triangle removal lemma in a restricted setting, when the triangles in the graph satisfy a certain pseudorandom property. First, an equivalent formulation of the triangle removal lemma is that any graph that contains εN^2 edge disjoint triangles, must contain at least δN^3 triangles. Note that we can identify a triangle (x,y,z) with a point in $[N]^3$. We may also assume without loss of generality that the graphs we are studying are tri-partite. Thus, we can identify the triangles in a graph G with a subset of $[N]^3$.

[3]That is, one needs to remove at least $\varepsilon\binom{N}{2}$ edges in order to make the graph triangle free.

Assume we are in the following special case. Let G be a tri-partite graph with N nodes on each side. Let $A \subset [N]^3$ denote a maximal collection of edge-disjoint triangles in G, and assume that $|A| = 2^{-d}N^2$. Observe that A is a weak permutation function. Assume that there exists a weak permutation function $S \subset [N]^3$ which contains A, and which is γ-pseudorandom against triangles for $\gamma \leqslant 2^{-O(d^4)}$. Theorem 2.4 then implies that G contains at least $2^{-O(d^3)}N^3$ many triangles. That is, in this special case we obtain $\delta \geqslant \varepsilon^{O(\log^2(1/\varepsilon))}$.

5 Obtaining a structured cube

In this section, we prove Lemma 2.7. To recall, the setup is the following: we have some weak permutation function $S \subset [N]^3$ which is pseudorandom against cubes, and we have a set $A \subset S$ which is dense in S. The goal is to perform a density increment process which restricts A to a large cube $C = X \times Y \times Z$ where A satisfies various pseudorandom properties. First, the cube C should be relatively dense in $[N]^3$. Second, the marginals of $A \cap C$ restricted to the XY, XZ, YZ-faces should be spread. As a technical condition, we will also want the rows of the marginals to be left lower-bounded.

While this third condition is seemingly harmless, we were unable to obtain left lower-boundedness using the traditional density increment process which iteratively restricts to better and better cubes. To circumvent this obstacle, we follow [KLM24] which optimizes a carefully defined potential function. As in their work, we first prove an intermediate result (Lemma 5.2) which only guarantees left lower-boundedness for a large fraction of the rows. We will later follow the proof of [KLM24, Lemma 2.13] to prune out these bad rows while maintaining the desired pseudorandom properties.

Definition 5.1 (Mostly left lower-bounded)**.** *Let $\varepsilon \in (0,1), \beta \in [0,1]$. A function $f : X \times Y \to [0,1]$ is* β-mostly ε-left lower-bounded *if for at least a $(1-\beta)$-fraction of $x \in X$, we have*

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

Lemma 5.2. *Let $S \subset [N]^3$ be a weak permutation function which is γ-pseudorandom against cubes. Let $d \geqslant 1, r \geqslant 1, \varepsilon \in (0,1), \beta \in (0,1/2)$, and assume $\gamma \leqslant 2^{-\Omega(dr/\varepsilon)}\beta$. Let $A \subset S$ of size $|A| \geqslant 2^{-d}|S|$. Then there is a cube $C = X \times Y \times Z \subset [N]^3$ of size $|C| \geqslant 2^{-O(dr/\varepsilon)}N^3$ such that*

1. *$|A \cap C| \geqslant 2^{-d}|S \cap C|$,*
2. *A_{XY}, A_{XZ}, A_{YZ} are (r,ε)-spread,*
3. *A_{XY}, A_{XZ}, A_{YZ} are β-mostly ε-left lower-bounded.*

Proof. We make minor modifications to the proof of [KLM24, Lemma 5.5]. Let $\eta = \Theta(\varepsilon/r)$ be sufficiently small. Given a cube $C \subset [N]^3$, define the potential function

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

Let $C = X \times Y \times Z$ be a cube which maximizes $\phi(\cdot)$.

Density and Large Cube. Initially for $C_0 = [N]^3$ we have $\phi(C_0) = (|A|/|S|)N^{3\eta} \geqslant 2^{-d}N^{3\eta}$, and for C we have $\phi(C) = (|A \cap C|/|S \cap C|)|C|^\eta$. Since C maximizes $\phi(\cdot)$ we can already make two deductions. First, since $|C| \leqslant N^3$ we must have $|A \cap C| \geqslant 2^{-d}|S \cap C|$, and second, since $|A \cap C| \leqslant |S \cap C|$ we have $|C|^\eta \geqslant 2^{-d}N^{3\eta}$, which implies $|C| \geqslant 2^{-d/\eta}N^3 \geqslant 2^{-O(dr/\varepsilon)}N^3$, with the last inequality holding for $\eta = \Omega(\varepsilon/r)$.

Spreadness. We will show that the marginal A_{XY} is (r, ε)-spread. Showing spreadness for A_{XZ} and A_{YZ} is similar. Assume towards a contradiction that there exists a rectangle $R = X' \times Y' \subset X \times Y$ of size $|R| \geqslant 2^{-r}|X||Y|$ such that

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

Define $C' = X' \times Y' \times Z$. We have

$$\mathop{\mathbb{E}}_{x \in X', y \in Y'} A_{XY}(x, y) = \frac{|A \cap C'|}{|X'||Y'|} \quad \text{and} \quad \mathop{\mathbb{E}}_{x \in X, y \in Y} A_{XY}(x, y) = \frac{|A \cap C|}{|X||Y|}$$

where we have critically used the property that S is a weak permutation function. This gives

$$\frac{|A \cap C'|}{|A \cap C|} > (1 + \varepsilon) \cdot \frac{|X'||Y'|}{|X||Y|} = (1 + \varepsilon) \cdot \frac{|C'|}{|C|}.$$

Let $|S| = \sigma N^3$, noting that $\sigma \leqslant 1/N$. By the γ-pseudorandomness of S against cubes, we have $|S \cap C| = \sigma|C| \pm \gamma|S|$ and $|S \cap C'| = \sigma|C'| \pm \gamma|S|$. Thus,

$$\frac{|S \cap C|}{|S \cap C'|} \geqslant \frac{\sigma|C| - \gamma|S|}{\sigma|C'| + \gamma|S|} \geqslant \frac{1 - \frac{\gamma|S|}{\sigma|C|}}{1 + \frac{\gamma|S|}{\sigma|C'|}} \cdot \frac{\sigma|C|}{\sigma|C'|} = \frac{1 - \gamma\frac{N^3}{|C|}}{1 + \gamma\frac{N^3}{|C'|}} \cdot \frac{|C|}{|C'|}.$$

Recall that $|C| \geqslant 2^{-O(dr/\varepsilon)}N^3$. Similarly, $|C'| \geqslant 2^{-r}|C| \geqslant 2^{-O(dr/\varepsilon)}N^3$. Therefore,

$$\frac{|S \cap C|}{|S \cap C'|} \geqslant \frac{1 - \varepsilon/16}{1 + \varepsilon/16} \cdot \frac{|C|}{|C'|} \geqslant (1 - \varepsilon/4)\frac{|C|}{|C'|}$$

where the penultimate inequality holds for $\gamma \leqslant 2^{-\Omega(dr/\varepsilon)}$ small enough. Putting everything together, we get

$$\frac{\phi(C')}{\phi(C)} = \frac{|A \cap C'|}{|A \cap C|} \cdot \frac{|S \cap C|}{|S \cap C'|} \cdot \left(\frac{|C'|}{|C|}\right)^{\eta} > (1 + \varepsilon)(1 - \varepsilon/4) \cdot 2^{-r\eta} > 1$$

where the last inequality holds for $\eta \leqslant O(\varepsilon/r)$ small enough. This contradicts the maximality of C.

Mostly left lower-bounded. We next show that the marginal A_{XY} is β-mostly ε-left lower-bounded. The remaining marginals follow similarly. Assume towards a contradiction that there exists $X' \subset X$ of size $|X'| = \beta|X|$ such that

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

Set $C' = X' \times Y \times Z$, and let $C'' = C \backslash C' = (X \backslash X') \times Y \times Z$. Note that $|C''| = (1 - \beta)|C|$. We will show that $\phi(C'') > \phi(C)$, which is a contradiction to the maximality of C. We have

$$|A \cap C''| = |A \cap C| - |A \cap C'| > (1 - (1 - \varepsilon)\beta)|A \cap C|.$$

At this point, we proceed similarly to the previous paragraph. By the γ-pseudorandomness of S, we have $|S \cap C| = \sigma|C| \pm \gamma|S|$ and $|S \cap C''| = \sigma|C''| \pm \gamma|S|$. This gives

$$\frac{|S \cap C|}{|S \cap C''|} \geqslant \frac{\sigma|C| - \gamma|S|}{\sigma|C''| + \gamma|S|} \geqslant \frac{1 - \gamma\frac{N^3}{|C|}}{1 + \gamma\frac{N^3}{|C''|}} \cdot \frac{\sigma|C|}{\sigma|C''|} \geqslant \frac{1 - \gamma \cdot 2^{O(dr/\varepsilon)}}{1 + \gamma \cdot 2^{O(dr/\varepsilon)}} \cdot \frac{|C|}{|C''|} \geqslant (1 - \gamma \cdot 2^{O(dr/\varepsilon)})(1 - \beta)^{-1}.$$

Putting everything together, we get

$$\frac{\phi(C'')}{\phi(C)} = \frac{|A \cap C''|}{|A \cap C|} \cdot \frac{|S \cap C|}{|S \cap C''|} \cdot \left(\frac{|C''|}{|C|}\right)^{\eta} > (1 - \beta + \varepsilon\beta)(1 - \gamma \cdot 2^{O(dr/\varepsilon)})(1 - \beta)^{-1}(1 - \beta)^{\eta} > 1$$

where the last inequality holds for $\gamma \leqslant \varepsilon\beta \cdot 2^{-\Omega(dr/\varepsilon)}$ and $\eta \leqslant O(\varepsilon)$ small enough. $\square$

At this point, we prune out the rows which are not left lower-bounded. Our proof is similar to [KLM24, Lemma 2.13].

Lemma 2.7 (Restricting to a good cube)**.** *Let $S \subset [N]^3$ be a weak permutation function which is γ-pseudorandom against cubes. Let $d \geqslant 1, r \geqslant 1, \varepsilon \in (0,1)$, and assume $\gamma \leqslant 2^{-\Omega(dr/\varepsilon)}$. Let $A \subset S$ of size $|A| \geqslant 2^{-d}|S|$. Then there is a cube $C = X \times Y \times Z \subset [N]^3$ of size $|C| \geqslant 2^{-O(dr/\varepsilon)}N^3$ with the following properties:*

1. $|A \cap C| \geqslant 2^{-(d+1)}|S \cap C|$,
2. A_{XZ}, A_{YZ} *are (r,ε)-spread,*
3. A_{XZ}, A_{YZ} *are ε-left lower-bounded.*

Proof. Apply Lemma 5.2 with parameters d, $r+1$, $\varepsilon/2$, $\beta = 2^{-\Theta(dr/\varepsilon)}$ small enough, which we can as we assume $\gamma \leqslant 2^{-\Omega(dr/\varepsilon)}$. We next prune $C = X \times Y \times Z$ to obtain the desired cube.

Let $X' \subset X$ be the set of points where A_{XZ} is $(\varepsilon/2)$-left lower-bounded, and $Y' \subset Y$ be the set of points y where A_{YZ} is $(\varepsilon/2)$-left lower-bounded, both with respect to C. Let $C' = X' \times Y' \times Z$, and consider the marginals $A_{X'Z}, A_{Y'Z}$ with respect to C'. The claim is that the cube C' satisfies the desired properties.

We already have $|A \cap C| \geqslant 2^{-d}|S \cap C|$, and we will now show that $|A \cap C'| \geqslant 2^{-(d+1)}|S \cap C'|$. We will do this by arguing that the number of points in $A \cap (C \backslash C')$ is small. Each $x \in X \backslash X'$ satisfies

$$\mathop{\mathbb{E}}_{z \in Z} A_{XZ}(x,z) \leqslant (1-\varepsilon/2)\,\mathbb{E}[A_{XZ}],$$

so the "bad" rows of A_{XZ} account for at most $(1-\varepsilon/2)\beta|A \cap C|$ many points, and similarly for A_{YZ}. This gives

$$|A \cap C'| \geqslant |A \cap C| - 2(1-\varepsilon/2)\beta|A \cap C| \geqslant (1-2\beta)|A \cap C|.$$

Combined with $|S \cap C'| \leqslant |S \cap C|$, and assuming $\beta \leqslant 1/4$, we obtain $|A \cap C'| \geqslant 2^{-(d+1)}|S \cap C'|$. Next, observe that

$$\mathbb{E}[A_{X'Z}] = \frac{|A \cap C'|}{|X'||Z|} \geqslant \frac{(1-2\beta)|A \cap C|}{|X||Z|} = (1-2\beta)\,\mathbb{E}[A_{XZ}],$$

and similarly

$$\mathbb{E}[A_{X'Z}] = \frac{|A \cap C'|}{|X'||Z|} \leqslant \frac{|A \cap C|}{(1-\beta)|X||Z|} = (1-\beta)^{-1}\,\mathbb{E}[A_{XZ}],$$

which will later be used to show left lower-boundedness.

We now show that $A_{X'Z}, A_{Y'Z}$ are (r,ε)-spread. We show this for $A_{X'Z}$, and an analogous argument works for $A_{Y'Z}$. Assume that $R \subset X' \times Z$ is a rectangle of size $|R| \geqslant 2^{-r}|X'||Z|$. We can also view R as a rectangle $R \subset X \times Z$ of size $|R| \geqslant (1-\beta)2^{-r}|X||Z| \geqslant 2^{-(r+1)}|X||Z|$. Note that $A_{X'Z}$ is upper bounded by A_{XZ}. That is, for $(x,z) \in X' \times Z$ we have

$$A_{X'Z}(x,z) = \mathbb{1}[\exists y \in Y', (x,y,z) \in A] \leqslant \mathbb{1}[\exists y \in Y, (x,y,z) \in A] = A_{XZ}(x,z).$$

Next, applying the assumption that A_{XZ} is $(r+1,\varepsilon/2)$-spread gives

$$\mathop{\mathbb{E}}_{(x,z)\in R} A_{X'Z}(x,z) \leqslant \mathop{\mathbb{E}}_{(x,z)\in R} A_{XZ}(x,z) \leqslant (1+\varepsilon/2)\,\mathbb{E}[A_{XZ}] \leqslant (1+\varepsilon/2)(1-2\beta)^{-1}\,\mathbb{E}[A_{X'Z}].$$

We may assume $\beta = O(\varepsilon)$ is small enough so that $(1+\varepsilon/2)(1-2\beta)^{-1} \leqslant 1+\varepsilon$, which concludes the proof of spreadness.

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

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

Next, since S is a weak permutation function, we have

$$\sum_{z\in Z} A_{X'Z}(x,z) = |A\cap(\{x\}\times Y'\times Z)| \geqslant |A\cap(\{x\}\times Y\times Z)| - |Y\backslash Y'| \geqslant \sum_{z\in Z} A_{XZ}(x,z) - \beta N.$$

If instead we average over Z, using the fact that $|Z| \geqslant 2^{-O(dr/\varepsilon)}N$ gives

$$\mathop{\mathbb{E}}_{z\in Z} A_{X'Z}(x,z) \geqslant \mathop{\mathbb{E}}_{z\in Z} A_{XZ}(x,z) - \beta\cdot 2^{O(dr/\varepsilon)}.$$

Setting $\beta = 2^{-\Omega(dr/\varepsilon)}$ small enough, we can obtain

$$\mathop{\mathbb{E}}_{z\in Z} A_{X'Z}(x,z) \geqslant (1-\varepsilon/4)\mathop{\mathbb{E}}_{z\in Z} A_{XZ}(x,z).$$

We already saw that $\mathbb{E}[A_{XZ}] \geqslant (1-\beta)\,\mathbb{E}[A_{X'Z}]$. Putting this all together gives

$$\mathop{\mathbb{E}}_{z\in Z} A_{X'Z}(x,z) \geqslant (1-\varepsilon/4)\mathop{\mathbb{E}}_{z\in Z} A_{XZ}(x,z) \geqslant (1-\varepsilon/4)(1-\varepsilon/2)\,\mathbb{E}[A_{XZ}] \geqslant (1-\varepsilon/4)(1-\varepsilon/2)(1-\beta)\,\mathbb{E}[A_{X'Z}].$$

For $\beta \leqslant O(\varepsilon)$ small enough, the right hand side is at least $(1-\varepsilon)\,\mathbb{E}[A_{X'Z}]$. This concludes the proof of left lower-boundedness. □

6 Density increment for sparse functions

In this section, we prove Lemma 2.13, restated below.

Lemma 2.13 (Relative version of [KLM24, Lemma 4.7])**.** *Let $f : X\times Y \to [0,1]$ be a function supported on a set $T \subset X\times Y$ of size $\tau|X||Y|$; suppose that $\|f\|_{1(\mu_T)} = \delta$. Let $\ell,k\in\mathbb{N}$. Additionally, assume that T is γ-pseudorandom with respect to rectangles for $\gamma \leqslant \varepsilon^8\cdot(\tau\delta/2)^{O(\ell k)}$ small enough. If*

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

then there exists some rectangle $R\subset X\times Y$ with

$$\mathop{\mathbb{E}}_{(x,y)\in R} f(x,y) \geqslant \left(1+\frac{\varepsilon}{64}\right)\|f\|_1 \quad \textit{and} \quad \|R\|_1 \geqslant \frac{1}{64}\cdot\varepsilon\cdot\delta^{\ell k+1}.$$

One may wish to compare it directly with the original version below, where the size of the rectangle obtained depends on the *global* density of f, rather than its density inside a sparse pseudorandom set.

Lemma 2.9 ([KLM24, Lemma 4.7])**.** *Let $f : X\times Y\to[0,1]$; suppose that $\|f\|_1 \geqslant \delta$. Let $\ell,k\in\mathbb{N}$. If*

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

then there exists some rectangle $R\subset X\times Y$ with

$$\mathop{\mathbb{E}}_{(x,y)\in R} f(x,y) \geqslant \left(1+\frac{\varepsilon}{2}\right)\|f\|_1 \quad \textit{and} \quad \|R\|_1 \geqslant \frac{1}{2}\cdot\varepsilon\cdot\delta^{\ell k+1}.$$

Much of our proof proceeds in the same way as the proof of Lemma 4.7 in [KLM24]. They begin with a function f defined on $X\times Y$ which is dense and has a large grid norm, and obtain a convex combination of rectangles which has high correlation with f. Then, they argue that if f has high correlation with a convex combination of rectangles, then it must admit a density increment on some large rectangle. More precisely, they show:

Claim 6.1 ([KLM24, Claim 4.6]). *Let $f\colon X \times Y \to \mathbb{R}_{\geqslant 0}$ and W be a convex combination of rectangles. Suppose that $\|f\|_\infty \leqslant \Delta$ and $\|W\|_1 \geqslant \delta$. If*

$$\left\langle \frac{W}{\|W\|_1}, f \right\rangle \geqslant 1 + \varepsilon,$$

then there is some rectangle R with

$$\left\langle \frac{R}{\|R\|_1}, f \right\rangle \geqslant 1 + \frac{\varepsilon}{2} \quad \text{and} \quad \|R\|_1 \geqslant \frac{\varepsilon\delta}{2\Delta}.$$

Our proof will follow the same outline. The major difference is that while their proof starts with a *globally* dense function, we begin with a function f which is dense on a sparse pseudorandom set T. The function f may be very sparse when viewed as a function on $X \times Y$, which is what prevents us from simply applying the claims from [KLM24]. Much of the work that goes into proving Lemma 2.13 is proving versions of Claim 4.6 and Lemma 4.7 from [KLM24] which work in this sparse setting. We reiterate that crucially, the density of the rectangles obtained by our analogous results will *not* depend on the density of T.

We will first prove a similar claim to Claim 6.1, but where the inner products and expectations in the conclusion are taken with respect to a distribution μ. For concreteness, we will eventually apply Claim 6.2 with the uniform distribution over a sparse pseudorandom set; however, the statement holds for arbitrary distributions. The proof follows that of [KLM24, Claim 4.6] with almost no modification.

Claim 6.2 (Relative version of [KLM24, Claim 4.6]). *Let $f : X \times Y \to \mathbb{R}_{\geqslant 0}$ and W be a convex combination of rectangles. Let μ be a distribution on $X \times Y$. Suppose that $\|f\|_\infty \leqslant \Delta$ and $\|W\|_{1(\mu)} \geqslant \delta$. If*

$$\left\langle \frac{W}{\|W\|_{1(\mu)}}, f \right\rangle_\mu \geqslant 1 + \varepsilon,$$

then there is some rectangle R with

$$\left\langle \frac{R}{\|R\|_{1(\mu)}}, f \right\rangle_\mu \geqslant 1 + \frac{\varepsilon}{2} \quad \text{and} \quad \|R\|_{1(\mu)} \geqslant \frac{\varepsilon\delta}{2\Delta}.$$

Proof. We follow the proof of [KLM24, Claim 4.6] and edit steps to work with distributions. Write $W = \sum_i c_i R_i$ where R_i are rectangles and $c_i \geqslant 0$, $\sum c_i = 1$. We begin by pruning rectangles which are too small; define $W' = \sum_i c_i' R_i$ via $c_i' = c_i$ if $\|R_i\|_{1(\mu)} \geqslant \kappa$ and 0 otherwise for some threshold value κ. We note that

$$\frac{\langle W', f\rangle_\mu}{\|W\|_{1(\mu)}} = \frac{\langle W, f\rangle_\mu}{\|W\|_{1(\mu)}} - \frac{\langle W - W', f\rangle_\mu}{\|W\|_{1(\mu)}} \geqslant 1 + \varepsilon - \frac{\|W - W'\|_{1(\mu)}\|f\|_\infty}{\|W\|_{1(\mu)}} \geqslant 1 + \varepsilon - \frac{\kappa\Delta}{\delta}.$$

Setting $\kappa = \varepsilon\delta/2\Delta$ gives

$$\frac{\langle W', f\rangle_\mu}{\|W'\|_{1(\mu)}} \geqslant \frac{\langle W', f\rangle_\mu}{\|W\|_{1(\mu)}} \geqslant 1 + \frac{\varepsilon}{2}.$$

In particular, we must have that $\langle W', f\rangle_\mu > 0$, which guarantees that W' is not identically zero on the support of μ. We have

$$\frac{\langle W', f\rangle_\mu}{\|W'\|_{1(\mu)}} = \frac{\sum_i c_i' \langle R_i, f\rangle_\mu}{\sum_i c_i' \|R_i\|_{1(\mu)}}.$$

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

$$\left\langle \frac{R}{\|R\|_{1(\mu)}}, f \right\rangle_\mu \geqslant 1 + \frac{\varepsilon}{2} \quad \text{and} \quad \|R\|_{1(\mu)} \geqslant \kappa. \qquad \square$$

We next prove a relative version of [KLM24, Lemma 4.7]. Our first adaptation requires pseudorandom properties of T slightly different from those we have used prior, but we will later show using Lemmas 3.3 and 6.7 that they follow from our more standard assumption of T being pseudorandom against rectangles.

The proof of Lemma 2.13 goes through a special distribution before moving to the uniform distribution on T. In order to define this distribution, we first require some new notation.

Definition 6.3 (Grid set)**.** *For a set $T \subset X \times Y$ and $\ell, k \in \mathbb{N}$, let the (ℓ,k)*-grid set *be $\Gamma_{\ell,k}(T) \subset X^\ell \times Y^k$ defined as*

$$\Gamma_{\ell,k}(T) := \{(x_1,\dots,x_\ell,y_1,\dots,y_k) : (x_i,y_j) \in T \text{ for every } i \in [\ell], j \in [k]\}.$$

The (ℓ,k)-grid set of T gives extra information in comparison to the (ℓ,k)-grid norm of T. It follows from the definition that

$$U_{\ell,k}(T) = \frac{|\Gamma_{\ell,k}(T)|}{|X|^\ell |Y|^k}.$$

Defining the (ℓ,k)-grid set has the added benefit of providing combinatorial information about T. The relationship between the (ℓ,k)-grid set and (ℓ,k)-grid norm will be the main focus in the proof of Lemma 6.6. It is helpful to think of the setting where T is the edge-set of a bipartite graph with vertex sets X and Y. For example, if the vertices $\{x_1,\dots,x_\ell\} \subset X$ and $\{y_1,\dots,y_k\} \subset Y$ form a $K_{\ell,k}$-minor, then $(\mathbf{x},\mathbf{y}) \in \Gamma_{\ell,k}(T)$. Note that there could be many more elements in the (ℓ,k)-grid set of T than $K_{\ell,k}$-minors since the tuples $(\mathbf{x},\mathbf{y})$ are ordered, and the entries are not required to be distinct. We will also want to define a set which contains the tuples $(\mathbf{x},\mathbf{y}) \in \Gamma_{\ell,k}(T)$ where $\mathbf{x}_1 = a, \mathbf{y}_1 = b$. This motivates the following definition:

Definition 6.4 (Restricted grid set)**.** *For a set $T \subset X \times Y$, $(a,b) \in X \times Y$, and $\ell, k \in \mathbb{N}$, let the (a,b)-restricted (ℓ,k)-grid set be*

$$\Gamma^{(a,b)}_{\ell,k}(T) := \{(\mathbf{x},\mathbf{y}) \in \Gamma_{\ell,k}(T) : \mathbf{x}_1 = a, \mathbf{y}_1 = b\}.$$

Note that we could have used any choice of indices $\mathbf{x}_i = a, \mathbf{y}_j = b$ in the definition, since the set $\Gamma_{\ell,k}(T)$ is invariant under permuting the first ℓ coordinates and the last k coordinates. If we return to the graph analogy, it is easy to see that every $K_{\ell,k}$-minor in the graph which contains the edge (a,b) will correspond to a tuple in $\Gamma^{(a,b)}_{\ell,k}(T)$. If Y_a, X_b denote the neighborhoods of a, b, respectively, the (a,b)-restricted (ℓ,k)-grid is equivalent to considering the set of $K_{\ell-1,k-1}$-minors contained in $X_b \times Y_a$. This perspective will be especially useful in the proof of Lemma 6.7.

Finally, we will define a distribution supported on T which captures the fraction of tuples $(\mathbf{x},\mathbf{y}) \in \Gamma_{\ell,k}(T)$ which have $\mathbf{x}_1 = a, \mathbf{y}_1 = b$.

Definition 6.5 (Restricted grid set distribution)**.** *For a set $T \subset X \times Y$, $(a,b) \in X \times Y$, and $\ell, k \in \mathbb{N}$, let the (a,b)-restricted (ℓ,k)-grid set distribution $\nu^T_{\ell,k}(a,b)$ be the marginal distribution of $(\mathbf{x}_1,\mathbf{y}_1)$ when picking uniform random $(\mathbf{x},\mathbf{y}) \in \Gamma_{\ell,k}(T)$. More precisely,*

$$\nu^T_{\ell,k}(a,b) := \frac{|\Gamma^{(a,b)}_{\ell,k}(T)|}{|\Gamma_{\ell,k}(T)|} \cdot |X||Y|.$$

The factor of $|X||Y|$ is placed to ensure that $\mathbb{E}[\nu^T_{\ell,k}] = 1$. In the graph analogy, this corresponds to a distribution on T which captures the fraction of $K_{\ell,k}$-minors containing a fixed edge (a,b). If the edge set T is chosen at random, we would expect that no edge is favored when it comes to being included in $K_{\ell,k}$-minors, and so $\nu^T_{\ell,k}$ should be close to uniform. In fact, it will suffice that T is pseudorandom against rectangles when viewed as a subset of $X \times Y$. We will later make this intuition precise in Lemma 6.7.

Now, we can state and prove Lemma 6.6.

Lemma 6.6. *Let $f : X \times Y \to [0,1]$ be a function supported on a set $T \subset X \times Y$ of size $\tau|X||Y|$; suppose that $\|f\|_{1(\mu_T)} = \delta$. Let $\ell, k \in \mathbb{N}$ and $\lambda \leqslant O(\varepsilon \cdot \delta^{\ell k+1})$ small enough. Additionally, assume that T satisfies the following pseudorandom properties:*

1. $\|T\|_{U(\ell,k)} \leqslant (1+\varepsilon/2)\tau$,

2. $\|\nu^T_{\ell,k} - \mu_T\|_1 \leqslant \lambda$.

If $\|f\|_{U(\ell,k)} \geqslant (1+\varepsilon)\|f\|_1$, *then there exists some rectangle* $R \subset X \times Y$ *with*

$$\left\langle \frac{R}{\|R\|_{1(\mu_T)}}, f \right\rangle_{\mu_T} \geqslant \left(1+\frac{\varepsilon}{32}\right)\|f\|_{1(\mu_T)} \quad \textit{and} \quad \|R\|_{1(\mu_T)} \geqslant \frac{1}{32}\cdot\varepsilon\cdot\delta^{\ell k+1}.$$

Proof. Assume $\delta > 0$, as otherwise the lemma trivially holds. For ease of notation, let $\mu = \mu_T$ and $\nu = \nu^T_{\ell,k}$. Since f is supported on T, we have $f = T \cdot f$. Additionally, $\|f\|_1 = \|T\|_1\|f\|_{1(\mu)}$, which implies $\|f\|_1 = \tau\delta$. By our assumption that $\|f\|_{U(\ell,k)} \geqslant (1+\varepsilon)\|f\|_1$, we have

$$\begin{aligned}
(1+\varepsilon)^{\ell k}(\tau\delta)^{\ell k} \leqslant \|f\|^{\ell k}_{U(\ell,k)} &= \mathop{\mathbb{E}}_{\substack{x_1,\ldots,x_\ell\in X\\ y_1,\ldots,y_k\in Y}}\left[\prod_{i=1}^{\ell}\prod_{j=1}^{k} f(x_i,y_j)\right]\\
&= \mathop{\mathbb{E}}_{\substack{x_1,\ldots,x_\ell\in X\\ y_1,\ldots,y_k\in Y}}\left[\left(\prod_{i=1}^{\ell}\prod_{j=1}^{k} T(x_i,y_j)\right)\left(\prod_{i=1}^{\ell}\prod_{j=1}^{k} f(x_i,y_j)\right)\right]\\
&= \Pr_{\substack{\mathbf{x}\in X^\ell\\ \mathbf{y}\in Y^k}}[(\mathbf{x},\mathbf{y})\in\Gamma_{\ell,k}(T)]\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y})\in\Gamma_{\ell,k}(T)}\left[\prod_{i=1}^{\ell}\prod_{j=1}^{k} f(\mathbf{x}_i,\mathbf{y}_j)\right].
\end{aligned}$$

A random $(\mathbf{x},\mathbf{y}) \in X^\ell \times Y^k$ is in $\Gamma_{\ell,k}(T)$ with probability at most $(1+\varepsilon/2)^{\ell k}\tau^{\ell k}$ by assumption (1), so

$$\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y})\in\Gamma_{\ell,k}(T)}\left[\prod_{i=1}^{\ell}\prod_{j=1}^{k} f(\mathbf{x}_i,\mathbf{y}_j)\right] \geqslant (1+\varepsilon)^{\ell k}(\tau\delta)^{\ell k}\cdot \Pr_{\substack{\mathbf{x}\in X^\ell\\ \mathbf{y}\in Y^k}}[(\mathbf{x},\mathbf{y})\in\Gamma_{\ell,k}(T)]^{-1} \geqslant \left(1+\frac{\varepsilon}{4}\right)^{\ell k}\delta^{\ell k}. \tag{1}$$

Now, we will use the same telescoping sum trick as in the proof of [KLM24, Lemma 4.7]. Fix some arbitrary ordering on tuples $(i,j) \in [\ell]\times[k]$ and consider the prefix-products

$$\phi_{\leqslant(i,j)}(\mathbf{x},\mathbf{y}) := \prod_{(i',j')\leqslant(i,j)} f(\mathbf{x}_{i'},\mathbf{y}_{j'}) \quad \text{and} \quad \phi_{<(i,j)}(\mathbf{x},\mathbf{y}) := \prod_{(i',j')<(i,j)} f(\mathbf{x}_{i'},\mathbf{y}_{j'})$$

with the convention $\phi_{<(1,1)}(\mathbf{x},\mathbf{y}) := 1$. For clarity, one should view $\phi_{\leqslant(i,j)}(\mathbf{x},\mathbf{y})$ and $\phi_{<(i,j)}(\mathbf{x},\mathbf{y})$ as functions on $X^\ell \times Y^k$. This way, we can apply our functions to points $(\mathbf{x},\mathbf{y}) \in \Gamma_{\ell,k}(T)$. Note that $\mathbb{E}_{\mathbf{x}\in X^\ell,\mathbf{y}\in Y^k}[\phi_{\leqslant(1,1)}(\mathbf{x},\mathbf{y})] = \|f\|_1$ and $\mathbb{E}_{\mathbf{x}\in X^\ell,\mathbf{y}\in Y^k}[\phi_{\leqslant(\ell,k)}(\mathbf{x},\mathbf{y})] = \|f\|^{\ell k}_{U(\ell,k)}$. Consider the telescoping product

$$\prod_{(i,j)\in[\ell]\times[k]}\frac{\mathbb{E}_{(\mathbf{x},\mathbf{y})\in\Gamma_{\ell,k}(T)}\,\phi_{\leqslant(i,j)}(\mathbf{x},\mathbf{y})}{\mathbb{E}_{(\mathbf{x},\mathbf{y})\in\Gamma_{\ell,k}(T)}\,\phi_{<(i,j)}(\mathbf{x},\mathbf{y})} = \mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y})\in\Gamma_{\ell,k}(T)}\left[\prod_{i=1}^{\ell}\prod_{j=1}^{k} f(\mathbf{x}_i,\mathbf{y}_j)\right].$$

By (1), this quantity is at least $\left(1+\frac{\varepsilon}{4}\right)^{\ell k}\delta^{\ell k}$, so we infer that for some choice of (i^*,j^*) we have

$$\frac{\mathbb{E}_{(\mathbf{x},\mathbf{y})\in\Gamma_{\ell,k}(T)}\,\phi_{\leqslant(i^*,j^*)}(\mathbf{x},\mathbf{y})}{\mathbb{E}_{(\mathbf{x},\mathbf{y})\in\Gamma_{\ell,k}(T)}\,\phi_{<(i^*,j^*)}(\mathbf{x},\mathbf{y})} \geqslant \left(1+\frac{\varepsilon}{4}\right)\delta. \tag{2}$$

We would now like to think of $\phi_{<(i^*,j^*)}(\mathbf{x},\mathbf{y})$ primarily as a function of $\mathbf{x}_{i^*}$ and $\mathbf{y}_{j^*}$. By an abuse of notation, we think of $\Gamma^{(a,b)}_{\ell,k}(T)$ as the set of $(\mathbf{x},\mathbf{y}) \in \Gamma_{\ell,k}(T)$ with $\mathbf{x}_{i^*} = a, \mathbf{y}_{j^*} = b$. Define the function

$$W(a,b) = \mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y})\in\Gamma^{(a,b)}_{\ell,k}(T)}\prod_{(i,j)<(i^*,j^*)} f(\mathbf{x}_i,\mathbf{y}_j).$$

We now rewrite the fraction in (2) in terms of W. For the numerator, we have

$$\begin{aligned}
\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y})\in\Gamma_{\ell,k}(T)} \phi_{\leqslant(i*,j*)}(\mathbf{x},\mathbf{y}) &= \mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y})\in\Gamma_{\ell,k}(T)} \prod_{(i,j)\leqslant(i*,j*)} f(\mathbf{x}_i,\mathbf{y}_j) \\
&= \mathop{\mathbb{E}}_{(a,b)\sim\nu} \left(\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y})\in\Gamma^{(a,b)}_{\ell,k}(T)} \prod_{(i,j)<(i*,j*)} f(\mathbf{x}_i,\mathbf{y}_j) \right) \cdot f(a,b) \\
&= \mathop{\mathbb{E}}_{(a,b)\sim\nu} W(a,b) f(a,b) \\
&= \langle W, f\rangle_\nu .
\end{aligned}$$

For the denominator, we have

$$\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y})\in\Gamma_{\ell,k}(T)} \phi_{<(i*,j*)}(\mathbf{x},\mathbf{y}) = \mathop{\mathbb{E}}_{(a,b)\sim\nu} \mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y})\in\Gamma^{(a,b)}_{\ell,k}(T)} \phi_{<(i*,j*)}(\mathbf{x},\mathbf{y}) = \mathop{\mathbb{E}}_{(a,b)\sim\nu} W(a,b) = \|W\|_{1(\nu)}.$$

At this point, we have established the following inequality:

$$\langle W, f\rangle_\nu \geqslant \left(1+\frac{\varepsilon}{4}\right)\delta\|W\|_{1(\nu)}. \tag{3}$$

We can additionally obtain a lower bound on $\|W\|_{1(\nu)}$:

$$\begin{aligned}
\|W\|_{1(\nu)} &= \mathop{\mathbb{E}}_{(a,b)\sim\nu} \mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y})\in\Gamma^{(a,b)}_{\ell,k}(T)} \prod_{(i,j)<(i*,j*)} f(\mathbf{x}_i,\mathbf{y}_j) \\
&= \mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y})\in\Gamma_{\ell,k}(T)} \prod_{(i,j)<(i*,j*)} f(\mathbf{x}_i,\mathbf{y}_j) \\
&\geqslant \mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y})\in\Gamma_{\ell,k}(T)} \left[\prod_{i=1}^{\ell}\prod_{j=1}^{k} f(\mathbf{x}_i,\mathbf{y}_j) \right] && \text{(since } \|f\|_\infty \leqslant 1) \\
&\geqslant \left(1+\frac{\varepsilon}{4}\right)^{\ell k} \delta^{\ell k}. && \text{(by (1))}
\end{aligned}$$

Our goal is to obtain the inequalities

$$\|W\|_{1(\mu)} \geqslant \delta^{\ell k} \quad \text{and} \quad \left\langle \frac{W}{\|W\|_{1(\mu)}}, \frac{f}{\|f\|_{1(\mu)}} \right\rangle_\mu \geqslant 1+\frac{\varepsilon}{16},$$

so that we may apply Claim 6.2 with W and $f' = f/\|f\|_{1(\mu)}$. Now, we use assumption (2) that $\|\nu-\mu\|_1 \leqslant \lambda \leqslant (\varepsilon/16)\cdot\delta^{\ell k+1}$ along with the fact that $\|W\|_\infty, \|f\|_\infty \leqslant 1$ to deduce

$$|\langle W, f\rangle_\nu - \langle W, f\rangle_\mu| \leqslant \lambda \quad \text{and} \quad |\|W\|_{1(\nu)} - \|W\|_{1(\mu)}| \leqslant \lambda. \tag{4}$$

To lower bound $\|W\|_{1(\mu)}$, it suffices to lower bound $\|W\|_{1(\nu)}$. This gives

$$\|W\|_{1(\mu)} \geqslant \|W\|_{1(\nu)} - \lambda \geqslant \left(1+\frac{\varepsilon}{4}\right)^{\ell k}\delta^{\ell k} - \frac{\varepsilon}{16}\cdot\delta^{\ell k+1} \geqslant \delta^{\ell k}, \tag{5}$$

Note that (5) implies $\lambda \leqslant (\varepsilon/16)\delta\|W\|_{1(\mu)}$. Thus,

$$\|W\|_{1(\nu)} \geqslant \|W\|_{1(\mu)} - \lambda \geqslant \left(1-\frac{\varepsilon}{16}\right)\|W\|_{1(\mu)}. \tag{6}$$

Combining (3), (4), and (6) we can lower bound the inner product by

$$\langle W, f\rangle_\mu \geqslant \langle W, f\rangle_\nu - \lambda \geqslant \left(1+\frac{\varepsilon}{4}\right)\left(1-\frac{\varepsilon}{16}\right)\delta\|W\|_{1(\mu)} - \frac{\varepsilon}{16}\cdot\delta\|W\|_{1(\mu)} \geqslant \left(1+\frac{\varepsilon}{16}\right)\delta\|W\|_{1(\mu)}.$$

At this point, we have

$$\left\langle \frac{W}{\|W\|_{1(\mu)}}, \frac{f}{\|f\|_{1(\mu)}} \right\rangle_\mu = \frac{\langle W, f\rangle_\mu}{\|W\|_{1(\mu)}\|f\|_{1(\mu)}} \geqslant 1 + \frac{\varepsilon}{16}.$$

Lastly, we will want to apply Claim 6.2. We have given a lower bound on $\|W\|_{1(\mu)}$ in (5), so it remains to argue that W is a convex combination of rectangles. We have

$$\prod_{(i,j)<(i^*,j^*)} f(\mathbf{x}_i, \mathbf{y}_j) = \left(\prod_{\substack{(i,j)<(i^*,j^*) \\ i\neq i^* \\ j \neq j^*}} f(\mathbf{x}_i, \mathbf{y}_j) \right) \left(\prod_{\substack{(i,j)<(i^*,j^*) \\ i=i^*}} f(\mathbf{x}_{i^*}, \mathbf{y}_j) \right) \left(\prod_{\substack{(i,j)<(i^*,j^*) \\ j=j^*}} f(\mathbf{x}_i, \mathbf{y}_{j^*}) \right).$$

For any fixing of the variables other than $\mathbf{x}_{i^*}$ and $\mathbf{y}_{j^*}$, each of the above factors depends on at most one of $\mathbf{x}_{i^*}$ or $\mathbf{y}_{j^*}$ but not both. Thus, W is a convex combination of soft rectangles, so Claim 3.2 implies it is also a convex combination of rectangles. As of now, we have following hypotheses:

1. W is a convex combination of rectangles
2. Setting $f' := f/\|f\|_{1(\mu)}$, we have $\|f'\|_\infty \leqslant 1/\delta$
3. $\|W\|_{1(\mu)} \geqslant \delta^{\ell k}$

Applying Claim 6.2 gives a rectangle R with

$$\left\langle \frac{R}{\|R\|_{1(\mu)}}, \frac{f}{\|f\|_{1(\mu)}} \right\rangle_\mu \geqslant 1 + \frac{\varepsilon}{32} \quad \text{and} \quad \|R\|_{1(\mu)} \geqslant \frac{1}{32} \cdot \varepsilon \cdot \delta^{\ell k+1}. \qquad \square$$

It remains to show the pseudorandom assumptions of Lemma 6.6 hold when T is pseudorandom against rectangles. The first condition is a consequence of Lemma 3.3, which says that if T is pseudorandom against rectangles, then T has bounded grid norms. The second condition will follow from Lemma 6.7, which says that if T is pseudorandom against rectangles, then $\nu^T_{\ell,k}$ and μ are close in L_1-distance.

For a moment, we return to the graph theoretic analogy. Consider a bipartite graph with vertex sets X, Y and edge set T. For a fixed edge $(a, b) \in T$, one can consider the fraction of $K_{\ell,k}$-minors which contain (a, b) as an edge. If T was chosen uniformly at random, each edge should participate in roughly the same number of $K_{\ell,k}$-minors. Equivalently, if the vertices $\{a, x_2, \ldots, x_\ell\} \subset X$ and $\{b, y_2, \ldots, y_k\} \subset Y$ form a $K_{\ell,k}$-minor, then $\{x_2, \ldots, x_\ell\} \subset X_b$ and $\{y_2, \ldots, y_k\} \subset Y_a$ form a $K_{\ell-1,k-1}$-minor. For most choices of (a, b), the restriction of T to $X_b \times Y_a$ should still look like a random graph, and so the number of $K_{\ell-1,k-1}$-minors will be concentrated around its expectation. In fact, we will show that T does not need to be picked uniformly at random; it suffices that T be pseudorandom against rectangles. In the graph setting, this is equivalent to saying that the edge density of any subgraph is within a small additive error of the global edge density. The following lemma formalizes this intuition.

Lemma 6.7. *Let $\ell, k \in \mathbb{N}$, and let $T \subset X \times Y$ be a set with size $|T| = \tau|X||Y|$ which is γ-pseudorandom against rectangles for $\gamma \leqslant (\tau/2)^{O(\ell k)}$ small enough. Then, we have $\|\nu^T_{\ell,k} - \mu_T\|_1 \leqslant O\left(\gamma^{1/8}\right)$.*

Before we begin the proof, we will need a lemma stating that if $T \subset X \times Y$ is pseudorandom against rectangles, then the row densities of T are concentrated around the mean. We will use the following lemma along the way.

Lemma 6.8 ([Gre05, Lemma 3.1])**.** *Consider a bipartite graph with vertex parts X and Y and edge density α. Let $d(x)$ denote the degree of a vertex x, and let $\varepsilon_1, \varepsilon_2 \in (0, 1)$. If there are at least $\varepsilon_1|X|$ vertices $x \in X$ such that $|d(x) - \alpha|Y|| > \varepsilon_2|Y|$ or at least $\varepsilon_2|Y|$ vertices $y \in Y$ such that $|d(y) - \alpha|X|| > \varepsilon_1|X|$, then there exist $X' \subset X$ and $Y' \subset Y$ with*

$$|X'| \geqslant \min\left(\frac{\varepsilon_1}{2}, \frac{\varepsilon_2}{2}\right)|X| \quad \textit{and} \quad |Y'| \geqslant \min\left(\frac{\varepsilon_1}{2}, \frac{\varepsilon_2}{2}\right)|Y|$$

such that the edge density in the subgraph induced by X' and Y' is at least $\alpha + \varepsilon_1\varepsilon_2/2$.

Lemma 6.9. *Let $T \subset X \times Y$ be a set which is γ-pseudorandom against rectangles. For a random $x \in X$, with probability at least $1 - 2\gamma^{1/4}$ we have*

$$\left| \mathop{\mathbb{E}}_{y \in Y} T(x,y) - \frac{|T|}{|X||Y|} \right| \leqslant 2\gamma^{1/4},$$

and similarly for a random $y \in Y$, with probability at least $1 - 2\gamma^{1/4}$ we have

$$\left| \mathop{\mathbb{E}}_{x \in X} T(x,y) - \frac{|T|}{|X||Y|} \right| \leqslant 2\gamma^{1/4}.$$

Proof. We will show the first conclusion holds; the proof for the second conclusion is analogous. Assume for the sake of contradiction that with probability at least $2\gamma^{1/4}$ over the choice of $x \in X$, we have

$$\left| \mathop{\mathbb{E}}_{y \in Y} T(x,y) - \frac{|T|}{|X||Y|} \right| > 2\gamma^{1/4}.$$

By Lemma 6.8, there exists a rectangle $R = X' \times Y' \subset X \times Y$ with

$$\mathop{\mathbb{E}}_{(x,y) \in R} T(x,y) \geqslant \frac{|T|}{|X||Y|} + 2\gamma^{1/2} \quad \text{and} \quad \frac{|R|}{|X||Y|} \geqslant \gamma^{1/2}.$$

This gives

$$\left| \mathop{\mathbb{E}}_{(x,y) \in T} [R(x,y)] - \mathop{\mathbb{E}}_{x \in X, y \in Y} [R(x,y)] \right| \geqslant \frac{|R|}{|T|} \cdot \left(\frac{|T|}{|X||Y|} + 2\gamma^{1/2} \right) - \frac{|R|}{|X||Y|} = 2\gamma^{1/2} \cdot \frac{|R|}{|T|} \geqslant 2\gamma,$$

which contradicts the γ-pseudorandomness of T. □

Proof of Lemma 6.7. For ease of notation, let $\mu = \mu_T$ and $\nu = \nu^T_{\ell,k}$. Recall that for $(a,b) \in X \times Y$,

$$\Gamma^{(a,b)}_{\ell,k}(T) = \{(\mathbf{x}, \mathbf{y}) \in \Gamma_{\ell,k}(T) : \mathbf{x}_1 = a, \mathbf{y}_1 = b\} \quad \text{and} \quad \nu(a,b) = \frac{|\Gamma^{(a,b)}_{\ell,k}(T)|}{|\Gamma_{\ell,k}(T)|} \cdot |X||Y|.$$

Note that if $(a,b) \notin T$, then $\nu(a,b) = 0$. We will compute $\nu(a,b)$ for some fixed $(a,b) \in T$. Define the neighborhoods of a, b to be

$$X_b := \{x \in X : (x,b) \in T\} \quad \text{and} \quad Y_a := \{y \in Y : (a,y) \in T\}.$$

Let $T_{a,b} = T \cap (X_b \times Y_a)$. The first observation is that

$$(x_1, \ldots, x_k, y_1, \ldots, y_k) \in \Gamma^{(a,b)}_{\ell,k}(T) \iff (x_2, \ldots, x_\ell, y_2, \ldots, y_k) \in \Gamma_{\ell-1,k-1}(T_{a,b}).$$

Thus, it suffices to estimate $\Gamma_{\ell-1,k-1}(T_{a,b})$. We have

$$\frac{\Gamma_{\ell-1,k-1}(T_{a,b})}{|X_b|^{\ell-1}|Y_a|^{k-1}} = \mathop{\mathbb{E}}_{\substack{x_1,\ldots,x_{\ell-1} \in X_b \\ y_1,\ldots,y_{k-1} \in Y_a}} \left[\prod_{i=1}^{\ell-1} \prod_{j=1}^{k-1} T_{a,b}(x_i, y_j) \right] = U_{\ell-1,k-1}(T_{a,b}). \tag{7}$$

Since T is γ-pseudorandom, we would expect that for typical a, b, we have $|X_b|, |Y_a|$ to be roughly $\tau|X|, \tau|Y|$, respectively. From there, to upper bound $\Gamma_{\ell-1,k-1}(T_{a,b})$ it suffices to upper bound $U_{\ell-1,k-1}(T_{a,b})$. If we knew that $T_{a,b}$ is pseudorandom against rectangles, we could apply Lemma 3.3, but a priori this may not be the case. Luckily, $T \subset X \times Y$ is γ-pseudorandom against rectangles, so restricting T to a large rectangle $X_b \times Y_a$ should result in a set which is also pseudorandom against rectangles with some loss in parameters. The loss will be small as long as $X_b \times Y_a$ is relatively large, which as previously noted should be around $\tau^2|X||Y|$ for typical (a,b). Combining these arguments will let us bound the upward deviations of ν from μ, ultimately providing the desired bound on $\|\nu - \mu\|_1$. The remainder of the proof is to nail down the exact quantitative details of these statements.

Typical (a,b) are good. By Lemma 6.9, with probability at least $1-4\gamma^{1/4}$ for a random $a\in X$ and $b\in Y$, we have

$$\left|\frac{|X_b|}{|X|}-\tau\right|\leqslant 2\gamma^{1/4}\quad\text{and}\quad\left|\frac{|Y_a|}{|Y|}-\tau\right|\leqslant 2\gamma^{1/4}.$$

We will call (a,b) *good* if these conditions are satisfied.

$T_{a,b}$ is pseudorandom against rectangles for good (a,b). We will show that $T_{a,b}=T\cap(X_b\times Y_a)$ is $\gamma_{a,b}$-pseudorandom with respect to rectangles when viewed as a subset of $X_b\times Y_a$, where

$$\gamma_{a,b}\leqslant 2\gamma\frac{|T|}{|T_{a,b}|}.$$

In particular, when (a,b) is good, we will show that $\gamma_{a,b}\leqslant 4\gamma\tau^{-2}$. Since T is γ-pseudorandom against rectangles, we have

$$\left|\frac{|T_{a,b}|}{|T|}-\frac{|X_b||Y_a|}{|X||Y|}\right|=\left|\mathop{\mathbb{E}}_{(x,y)\in T}[(X_b\times Y_a)(x,y)]-\mathop{\mathbb{E}}_{x\in X,y\in Y}[(X_b\times Y_a)(x,y)]\right|\leqslant\gamma.$$

Consider a rectangle $R\subset X_b\times Y_a$. We have

$$\mathop{\mathbb{E}}_{(x,y)\in T}R(x,y)=\frac{|T_{a,b}|}{|T|}\cdot\mathop{\mathbb{E}}_{(x,y)\in T_{a,b}}R(x,y)$$

and

$$\mathop{\mathbb{E}}_{x\in X,y\in Y}R(x,y)=\frac{|X_b|}{|X|}\cdot\frac{|Y_a|}{|Y|}\cdot\mathop{\mathbb{E}}_{x\in X_b,y\in Y_a}R(x,y).$$

Combining these two equations with an application of the triangle inequality gives

$$\begin{aligned}\frac{|T_{a,b}|}{|T|}\cdot\left|\mathop{\mathbb{E}}_{(x,y)\in T_{a,b}}[R(x,y)]-\mathop{\mathbb{E}}_{x\in X_b,y\in Y_a}[R(x,y)]\right|&\leqslant\left|\mathop{\mathbb{E}}_{(x,y)\in T}[R(x,y)]-\frac{|X_b|}{|X|}\cdot\frac{|Y_a|}{|Y|}\cdot\mathop{\mathbb{E}}_{x\in X_b,y\in Y_a}[R(x,y)]\right|\\&\quad+\left|\left(\frac{|X_b|}{|X|}\cdot\frac{|Y_a|}{|Y|}-\frac{|T_{a,b}|}{|T|}\right)\cdot\mathop{\mathbb{E}}_{x\in X_b,y\in Y_a}[R(x,y)]\right|\\&\leqslant\left|\mathop{\mathbb{E}}_{(x,y)\in T}[R(x,y)]-\mathop{\mathbb{E}}_{x\in X,y\in Y}[R(x,y)]\right|+\gamma\\&\leqslant 2\gamma\end{aligned}$$

where the last inequality follows by γ-pseudorandomness of T. Rearranging shows that $T_{a,b}\subset X_b\times Y_a$ is $2\gamma\frac{|T|}{|T_{a,b}|}$-pseudorandom against rectangles. At this point, we will bound $\gamma_{a,b}$ for good (a,b). By definition, good (a,b) satisfy

$$|X_b|=(\tau\pm 2\gamma^{1/4})|X|\quad\text{and}\quad|Y_a|=(\tau\pm 2\gamma^{1/4})|Y|.$$

The γ-pseudorandomness of T gives

$$\frac{|T_{a,b}|}{|T|}\geqslant\frac{|X_b||Y_a|}{|X||Y|}-\gamma\geqslant(\tau-2\gamma^{1/4})^2-\gamma\geqslant\tau^2/2\tag{8}$$

where the last inequality holds for $\gamma\leqslant(\tau/2)^{O(1)}$ small enough. Taking inverses of both sides gives

$$\gamma_{a,b}\leqslant 2\gamma\frac{|T|}{|T_{a,b}|}\leqslant 4\gamma\tau^{-2},$$

and so $T_{a,b}\subset X_b\times Y_a$ is $4\gamma\tau^{-2}$-pseudorandom against rectangles for good (a,b).

$T_{a,b}$ has small grid norms for good (a, b). The above paragraph showed that $T_{a,b} \subset X_b \times Y_a$ is $4\gamma\tau^{-2}$-pseudorandom against rectangles for good (a, b). By Lemma 3.3, we know that

$$U_{\ell-1,k-1}(T_{a,b}) \leqslant (1+\gamma^{1/4})^{(\ell-1)(k-1)} \left(\frac{|T_{a,b}|}{|X_b||Y_a|}\right)^{(\ell-1)(k-1)} \tag{9}$$

as long as

$$4\gamma\tau^{-2} < \frac{1}{4} \cdot \gamma^{1/2} \cdot \left(\frac{|T_{a,b}|}{|X_b||Y_a|}\right)^{(\ell-1)(k-1)+1}.$$

We now verify the latter inequality. By the γ-pseudorandomness of T, we have

$$|T_{a,b}| = \tau|X_b||Y_a| \pm \gamma|T|,$$

and for good (a, b), we have

$$|X_b| = (\tau \pm 2\gamma^{1/4})|X| \quad \text{and} \quad |Y_a| = (\tau \pm 2\gamma^{1/4})|Y|.$$

We showed in (8) that $\frac{|T_{a,b}|}{|T|} \geqslant \tau^2/2$ for good (a, b). Similarly,

$$\frac{|T_{a,b}|}{|X_b||Y_a|} \geqslant \frac{\tau|X_b||Y_a| - \gamma|T|}{|X_b||Y_a|} \geqslant \tau - \frac{\gamma|T|}{|X_b||Y_a|} \geqslant \tau - \frac{\gamma|T|}{(\tau - 2\gamma^{1/4})^2|X||Y|} = \tau - \frac{\gamma\tau}{(\tau - 2\gamma^{1/4})^2} \geqslant \tau/2$$

where the last inequality holds for $\gamma \leqslant (\tau/2)^{O(1)}$ small enough. Thus,

$$\frac{1}{4} \cdot \gamma^{1/2} \cdot \left(\frac{|T_{a,b}|}{|X_b||Y_a|}\right)^{(\ell-1)(k-1)+1} \geqslant \frac{1}{4} \cdot \gamma^{1/2} \cdot \left(\frac{\tau}{2}\right)^{(\ell-1)(k-1)+1} \geqslant 4\gamma\tau^{-2},$$

where the last inequality holds for $\gamma \leqslant (\tau/2)^{O(\ell k)}$ small enough.

Bounding upward deviations of ν. At this point, for good (a, b), we have established the inequality

$$\begin{aligned}
\Gamma^{(a,b)}_{\ell,k}(T) &= U_{\ell-1,k-1}(T_{a,b}) \cdot |X_b|^{\ell-1}|Y_a|^{k-1} && \text{(by (7))} \\
&\leqslant (1+\gamma^{1/4})^{(\ell-1)(k-1)} \left(\frac{|T_{a,b}|}{|X_b||Y_a|}\right)^{(\ell-1)(k-1)} \cdot |X_b|^{\ell-1}|Y_a|^{k-1} && \text{(by (9))} \\
&\leqslant \left(1+\gamma^{1/8}\right) \left(\frac{|T_{a,b}|}{|X_b||Y_a|}\right)^{(\ell-1)(k-1)} \cdot |X_b|^{\ell-1}|Y_a|^{k-1} && (10)
\end{aligned}$$

where the last inequality holds for $\gamma \leqslant 2^{-O(\ell k)}$ small enough, using the fact that $(1 + x)^r \leqslant 1 + 2^r x$ for $x \in [0, 1]$ and $r \geqslant 1$. We will bound the latter two factors of (10) separately. Similar to the above paragraph, we have

$$\frac{|T_{a,b}|}{|X_b||Y_a|} \leqslant \frac{\tau|X_b||Y_a| + \gamma|T|}{|X_b||Y_a|} \leqslant \tau + \frac{\gamma|T|}{|X_b||Y_a|} \leqslant \tau + \frac{\gamma\tau}{(\tau - 2\gamma^{1/4})^2} \leqslant \left(1+\gamma^{1/4}\right)\tau$$

where the last inequality holds for $\gamma \leqslant (\tau/2)^{O(1)}$ small enough. We also have

$$|X_b|^{\ell-1}|Y_a|^{k-1} \leqslant \left(\tau + 2\gamma^{1/4}\right)^{\ell+k-2}|X|^{\ell-1}|Y|^{k-1} \leqslant \left(1+\gamma^{1/8}\right)\tau^{\ell+k-2}|X|^{\ell-1}|Y|^{k-1}$$

where the last inequality holds for $\gamma \leqslant 2^{-O(\ell+k)} \cdot \tau^{O(1)}$ small enough, again using the fact that $(1+x)^r \leqslant 1+2^r x$ for $x \in [0, 1]$ and $r \geqslant 1$. Combining the above three inequalities we obtain

$$\Gamma^{(a,b)}_{\ell,k}(T) \leqslant \left(1+\gamma^{1/8}\right)\left(\left(1+\gamma^{1/4}\right)\tau\right)^{(\ell-1)(k-1)} \cdot \left(1+\gamma^{1/8}\right)\tau^{\ell+k-2}|X|^{\ell-1}|Y|^{k-1}$$

$$\leqslant \left(1+\gamma^{1/8}\right)^3 \tau^{\ell k-1}|X|^{\ell-1}|Y|^{k-1},$$

with yet again the last inequality holding for $\gamma \leqslant 2^{-O(\ell k)} \cdot \tau^{O(1)}$ small enough. On the other hand, by monotonicity of grid norms (Claim 3.4), we have

$$|\Gamma_{\ell,k}(T)| = U_{\ell,k}(T) \cdot |X|^\ell |Y|^k \geqslant U_{1,1}(T)^{\ell k} \cdot |X|^\ell |Y|^k = \tau^{\ell k}|X|^\ell |Y|^k.$$

This gives

$$\nu(a,b) = \frac{|\Gamma^{(a,b)}_{\ell,k}(T)|}{|\Gamma_{\ell,k}(T)|} \cdot |X||Y| \leqslant \frac{\left(1+\gamma^{1/8}\right)^3 \tau^{\ell k-1}|X|^\ell |Y|^k}{\tau^{\ell k}|X|^\ell |Y|^k} \leqslant \frac{\left(1+\gamma^{1/8}\right)^3}{\tau} \leqslant \frac{\left(1+O(\gamma^{1/8})\right)|X||Y|}{|T|}.$$

We want to bound $\|\nu-\mu\|_1$. At this point, it is helpful to recall that μ is the uniform distribution over T, so $\mu(x,y) = |X||Y|/|T|$ for $(x,y) \in T$ and 0 otherwise. Additionally, remember that if a pair $(x,y) \in X \times Y$ is good, it must be in T. Since $\mathbb{E}[\nu], \mathbb{E}[\mu] = 1$, it suffices to bound the upward deviations of $\nu - \mu$, as this is within a factor of 2 of $\|\nu-\mu\|_1$. With probability at least $1 - 4\gamma^{1/4}$, we pick a good pair (x,y) which gives

$$\frac{1}{2} \cdot \|\nu-\mu\|_1 = \mathop{\mathbb{E}}_{x\in X, y\in Y}[\mathbb{1}_{\nu \geqslant \mu} \cdot (\nu-\mu)] \leqslant \mathop{\mathbb{E}}_{x\in X, y\in Y}\left[\mathbb{1}_{\nu\geqslant\mu} \cdot \mathbb{1}_{(x,y) \text{ is good}} \cdot \left(\nu - \frac{|X||Y|}{|T|}\right)\right] + 4\gamma^{1/4} \leqslant O(\gamma^{1/8}). \quad \square$$

To finish the proof of Lemma 2.13, we will combine Lemma 3.3 and Lemma 6.7 to argue that if T is pseudorandom against rectangles, then T satisfies the conditions of Lemma 6.6. The conclusion of Lemma 6.6 gives a large rectangle R so that (1) f admits a density increment under μ_T when restricted to R and (2) R is dense in T. To finish the proof, we use the fact that T is pseudorandom against rectangles to argue that R must also be dense globally. From there, we can turn the density increment obtained under μ_T to a density increment under the uniform distribution on the entire space. Critically, the density of R in T, and therefore the density of R in $X \times Y$, will not depend on the density of T in $X \times Y$.

Proof of Lemma 2.13. For ease of notation, let $\mu = \mu_T$ and $\nu = \nu^T_{\ell,k}$. For $\gamma \leqslant \varepsilon^2 \cdot \tau^{O(\ell k)}$ small enough, we can apply Lemma 3.3 and Lemma 6.7 to deduce that

1. $\|T\|_{U(\ell,k)} \leqslant (1+\varepsilon/2)\,\mathbb{E}[T]$,
2. $\|\nu-\mu\|_1 \leqslant O(\gamma^{1/8})$.

Since the conditions of Lemma 6.6 are satisfied for $\gamma \leqslant \varepsilon^8 \cdot (\delta/2)^{O(\ell k)}$ sufficiently small, we obtain a rectangle R with

$$\mathop{\mathbb{E}}_{(x,y)\sim\mu}\left[\frac{R(x,y)}{\|R\|_{1(\mu)}} f(x,y)\right] = \left\langle \frac{R}{\|R\|_{1(\mu)}}, f\right\rangle_\mu \geqslant \left(1+\frac{\varepsilon}{32}\right)\|f\|_{1(\mu)} \tag{11}$$

and $\|R\|_{1(\mu)} \geqslant \frac{1}{32} \cdot \varepsilon \cdot \delta^{\ell k+1}$. Note that by γ-pseudorandomness of T, we have

$$\|R\|_1 \geqslant \|R\|_{1(\mu)} - \gamma \geqslant \frac{1}{64} \cdot \varepsilon \cdot \delta^{\ell k+1} \tag{12}$$

for $\gamma \leqslant O(\varepsilon \cdot \delta^{\ell k+1})$ small enough. Rearranging terms, we obtain

$$\begin{aligned}
\mathop{\mathbb{E}}_{(x,y)\sim\mu}\left[\frac{R(x,y)}{\|R\|_{1(\mu)}} f(x,y)\right] &= \mathop{\mathbb{E}}_{x\in X, y\in Y}\left[\frac{R(x,y)}{\|R\|_{1(\mu)}} f(x,y)\mu(x,y)\right] \\
&= \frac{\|R\|_1}{\|R\|_{1(\mu)}} \mathop{\mathbb{E}}_{(x,y)\in R}[f(x,y)\mu(x,y)] \\
&= \frac{\|R\|_1}{\tau\|R\|_{1(\mu)}} \mathop{\mathbb{E}}_{(x,y)\in R}[f(x,y)]
\end{aligned}$$

where the last equality uses the fact that f is supported on T and $\mu(x,y) = \tau^{-1} \cdot T(x,y)$. Combining with (11) gives

$$\mathop{\mathbb{E}}_{(x,y)\in R} f(x,y) \geqslant \left(1+\frac{\varepsilon}{32}\right) \cdot \frac{\tau\|R\|_{1(\mu)}}{\|R\|_1} \cdot \|f\|_{1(\mu)}. \tag{13}$$

Since T is γ-pseudorandom, we have

$$\frac{\tau\|R\|_{1(\mu)}}{\|R\|_1} = \frac{|R\cap T|}{|R|} \geqslant \tau - \frac{\gamma|T|}{|R|} = \left(1-\frac{\gamma|T|}{\tau|R|}\right)\tau = \left(1-\frac{\gamma}{\|R\|_1}\right)\tau \geqslant \left(1-\frac{\varepsilon}{128}\right)\tau.$$

where the last inequality uses (12) and holds for $\gamma \leqslant O(\varepsilon^2 \cdot \delta^{\ell k+1})$ small enough. Finally substituting into (13) yields

$$\mathop{\mathbb{E}}_{(x,y)\in R} f(x,y) \geqslant \left(1+\frac{\varepsilon}{32}\right)\left(1-\frac{\varepsilon}{128}\right)\tau\|f\|_{1(\mu)} \geqslant \left(1+\frac{\varepsilon}{64}\right)\|f\|_1. \qquad \square$$

7 Sparse mixing

In this section, we prove Lemma 2.15. Our main tool will be the following lemma from [KLM24].

Lemma 7.1 ([KLM24, Lemma 4.8]). *Fix an even integer $p \in \mathbb{N}$, $\varepsilon \in (0,1/20)$, and set $k = \lceil p/\varepsilon \rceil$. Let $g : X \times Z \to \mathbb{R}_{\geqslant 0}$, $h : Y \times Z \to \mathbb{R}_{\geqslant 0}$ be two (nonzero) functions, and suppose that*

1. *$\|g\|_{U(2,k)} \leqslant (1+\varepsilon)\|g\|_1$,*
2. *$\|h\|_{U(2,k)} \leqslant (1+\varepsilon)\|h\|_1$,*
3. *g, h are ε-left lower-bounded.*

Then

$$\left\| \mathop{\mathbb{E}}_{z\in Z}[g(x,z)h(y,z)] - \mathbb{E}[g]\,\mathbb{E}[h] \right\|_p \leqslant 20\varepsilon\, \mathbb{E}[g]\,\mathbb{E}[h].$$

Lemma 2.15 (Sparse von Neumann). *Let $T \subset X \times Y$ be a set which is γ-pseudorandom against rectangles, and let $A \subset T$ be a set of size $|A| \geqslant 2^{-d}|T|$. Let $g : X \times Z \to [0,1], h : Y \times Z \to [0,1]$ be functions. Let $d \geqslant 1, \varepsilon \in (0,1/20)$. For $k = O(d/\varepsilon)$ a large enough integer, suppose that*

1. *$\|g\|_{U(2,k)} \leqslant (1+\varepsilon)\|g\|_1$,*
2. *$\|h\|_{U(2,k)} \leqslant (1+\varepsilon)\|h\|_1$,*
3. *g, h are ε-left lower-bounded,*
4. *$\gamma \leqslant (\varepsilon\|g\|_1\|h\|_1)^{O(d)}$ is small enough.*

Then

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

Proof. Assume g and h are nonzero, as otherwise the result trivially holds. Let $|T| = \tau|X||Y|$ and $D(x,y) = \mathbb{E}_{z\in Z}\, g(x,z)h(y,z) - \mathbb{E}[g]\,\mathbb{E}[h]$, and set $p = O(d)$ to be a large enough even integer. By Hölder's inequality, we have

$$\begin{aligned}\left| \mathop{\mathbb{E}}_{x\in X, y\in Y, z\in Z}[A(x,y)g(x,z)h(y,z)] - \mathbb{E}[A]\,\mathbb{E}[g]\,\mathbb{E}[h] \right| &= \left| \mathop{\mathbb{E}}_{x\in X, y\in Y} A(x,y)D(x,y) \right| \\ &= \tau \left| \mathop{\mathbb{E}}_{(x,y)\in T} A(x,y)D(x,y) \right| \end{aligned}$$

$$
\begin{aligned}
&\leqslant \tau \left(\mathop{\mathbb{E}}_{(x,y)\in T} A(x,y)^{p/(p-1)} \right)^{1-1/p} \left(\mathop{\mathbb{E}}_{(x,y)\in T} D(x,y)^p \right)^{1/p} \\
&\leqslant \tau \left(\frac{|A|}{|T|} \right)^{1-1/p} \left(\mathop{\mathbb{E}}_{(x,y)\in T} D(x,y)^p \right)^{1/p} \\
&\leqslant 2\,\mathbb{E}[A] \left(\mathop{\mathbb{E}}_{(x,y)\in T} D(x,y)^p \right)^{1/p} .
\end{aligned}
$$

Thus, it suffices to bound the p-norm of the function D restricted to the set T. We will argue that this quantity is within a small additive factor of $\|D\|_p$. Observe that for a fixed $z \in Z$, the function $g(x,z)h(y,z)$ is a soft rectangle. Trivially, $\mathbb{E}[g]\,\mathbb{E}[h]$ is also a soft rectangle. By Claim 3.2, we can write $D(x,y) = \sum_i c_i R_i$ where R_i are rectangles and $\sum_i |c_i| \leqslant 2$. The product of rectangles is also a rectangle, which means we can write D^p as a linear combination of rectangles where the coefficients' magnitudes sum to at most 2^p. Using the assumption that T is γ-pseudorandom against rectangles for $\gamma \leqslant (\varepsilon\|g\|_1\|h\|_1)^p$, we have

$$
\left(\mathop{\mathbb{E}}_{(x,y)\in T} D(x,y)^p \right)^{1/p} \leqslant \left(\mathop{\mathbb{E}}_{x\in X, y\in Y} D(x,y)^p + 2^p\gamma \right)^{1/p} \leqslant \|D\|_p + 2\gamma^{1/p} \leqslant \|D\|_p + 2\varepsilon\,\mathbb{E}[g]\,\mathbb{E}[h],
$$

where the penultimate inequality follows from concavity. We conclude by applying Lemma 7.1 to obtain $\|D\|_p \leqslant 20\varepsilon\,\mathbb{E}[g]\,\mathbb{E}[h]$. □

References

[Alo02] Noga Alon. Testing subgraphs in large graphs. *Random Structures & Algorithms*, 21(3-4):359–370, 2002.

[AS74] Miklós Ajtai and Endre Szemerédi. Sets of lattice points that form no squares. *Stud. Sci. Math. Hungar*, 9(1975):9–11, 1974.

[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.

[Aus16] Tim Austin. Ajtai–Szemerédi theorems over quasirandom groups. *Recent trends in combinatorics*, pages 453–484, 2016.

[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.

[Beh46] Felix A Behrend. On sets of integers which contain no three terms in arithmetical progression. *Proceedings of the National Academy of Sciences*, 32(12):331–332, 1946.

[BG08] Jean Bourgain and Alex Gamburd. Uniform expansion bounds for Cayley graphs of $SL_2(\mathbb{F}_p)$. *Annals of Mathematics*, pages 625–642, 2008.

[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.

[BK21] Amey Bhangale and Subhash Khot. Optimal inapproximability of satisfiable k-LIN over non-abelian groups. STOC 2021, page 1615–1628, New York, NY, USA, 2021. Association for Computing Machinery.

[BMZ97] Vitaly Bergelson, Randall McCutcheon, and Qing Zhang. A Roth theorem for amenable groups. *American Journal of Mathematics*, 119(6):1173–1211, 1997.

[BNP08] László Babai, Nikolay Nikolov, and László Pyber. Product growth and mixing in finite groups. In Shang-Hua Teng, editor, *Proceedings of the 19th Annual Symposium on Discrete Algorithms (SODA)*, pages 248–257, 2008.

[BNS89] L. Babai, N. Nisan, and M. Szegedy. Multiparty protocols and logspace-hard pseudorandom sequences. In *Proceedings of the Twenty-First Annual ACM Symposium on Theory of Computing*, STOC '89, page 1–11, New York, NY, USA, 1989. Association for Computing Machinery.

[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.

[CFZ14] David Conlon, Jacob Fox, and Yufei Zhao. Extremal results in sparse pseudorandom graphs. *Advances in Mathematics*, 256:206–290, 2014.

[CFZ15] David Conlon, Jacob Fox, and Yufei Zhao. A relative Szemerédi theorem. *Geometric and Functional Analysis*, 25(3):733–762, 2015.

[CG16] David Conlon and William Timothy Gowers. Combinatorial theorems in sparse random sets. *Annals of Mathematics*, pages 367–454, 2016.

[DLV24] Harm Derksen, Chin Ho Lee, and Emanuele Viola. Boosting uniformity in quasirandom groups: fast and simple. In *Proceedings of the 65th Symposium on Foundations of Computer Science (FOCS) (To appear). arXiv preprint arXiv:2409.06932*, 2024.

[FHHK24] Yuval Filmus, Hamed Hatami, Kaave Hosseini, and Esty Kelman. Sparse graph counting and Kelley-Meka bounds for binary systems. In *Proceedings of the 65th Symposium on Foundations of Computer Science (FOCS) (To appear). arXiv preprint arXiv:2311.12248*, 2024.

[Fox11] Jacob Fox. A new proof of the graph removal lemma. *Annals of Mathematics*, pages 561–579, 2011.

[Fro68] Ferdinand Georg Frobenius. Über gruppencharaktere (1896). *Gesammelte Abhandlungen III. Springer, Berlin–Heidelberg*, 1968.

[Gow08] William T Gowers. Quasirandom groups. *Combinatorics, Probability and Computing*, 17(3):363–387, 2008.

[Gre04] Ben Green. Finite field models in additive combinatorics. *arXiv preprint math/0409420*, 2004.

[Gre05] Ben Green. An argument of Shkredov in the finite field setting. *Preprint*, 2005.

[Gre21] Ben Green. Lower bounds for corner-free sets. *New Zealand Journal of Mathematics*, 51:1–2, Jul. 2021.

[GT08] Ben Green and Terence Tao. The primes contain arbitrarily long arithmetic progressions. *Annals of mathematics*, pages 481–547, 2008.

[GV19] W. T. Gowers and Emanuele Viola. Interleaved group products. *SIAM J. Comput.*, 48(2):554–580, 2019.

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

[Hun22] Zach Hunter. Corner-free sets via the torus. *arXiv preprint arXiv:2209.10012*, 2022.

[JLO24] Michael Jaber, Shachar Lovett, and Anthony Ostuni. Strong bounds for skew corner-free sets. *arXiv preprint arXiv:2404.07380*, 2024.

[KLM24] Zander Kelley, Shachar Lovett, and Raghu Meka. Explicit separations between randomized and deterministic Number-on-Forehead communication. In *Proceedings of the 56th Annual Symposium on Theory of Computing (STOC)*, pages 1299–1310, 2024.

[KM23] Zander Kelley and Raghu Meka. Strong bounds for 3-progressions. In *Proceedings of the 64th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 933–973, 2023.

[KRSS10] Yoshiharu Kohayakawa, Vojtěch Rödl, Mathias Schacht, and Jozef Skokan. On the triangle removal lemma for subgraphs of sparse pseudorandom graphs. In *An Irregular Mind: Szemerédi is 70*, pages 359–404. Springer, 2010.

[LL14] Nathan Linial and Zur Luria. An upper bound on the number of high-dimensional permutations. *Combinatorica*, 34(4):471–486, 2014.

[LM05] Michael T Lacey and William McClain. On an argument of Shkredov on two-dimensional corners. *arXiv preprint math/0510491*, 2005.

[LPS88] Alexander Lubotzky, Ralph Phillips, and Peter Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.

[LPS19] Nati Linial, Toniann Pitassi, and Adi Shraibman. On the communication complexity of high-dimensional permutations. In *Proceedings of the 10th Innovations in Theoretical Computer Science (ITCS)*, 2019.

[LS21a] Nati Linial and Adi Shraibman. An improved protocol for the Exactly-N problem. In Valentine Kabanets, editor, *36th Computational Complexity Conference, CCC 2021, July 20-23, 2021, Toronto, Ontario, Canada (Virtual Conference)*, volume 200 of *LIPIcs*, pages 2:1–2:8. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021.

[LS21b] Nati Linial and Adi Shraibman. Larger corner-free sets from better NOF Exactly-N protocols. *arXiv preprint arXiv:2102.00421*, 2021.

[Mek23] Raghu Meka. Strong bounds for 3-progressions, 2023. Talk as part the Breakthroughs lecture series at the Simons Institute for the Theory of Computing. [Accessed: 2024 03 16] `https://www.youtube.com/watch?v=WN7rJPWy6z8`.

[Mil24] Luka Milićević. Good bounds for sets lacking skew corners. *arXiv preprint arXiv:2404.07180*, 2024.

[NW91] Noam Nisan and Avi Widgerson. Rounds in communication complexity revisited. In *Proceedings of the twenty-third annual ACM symposium on Theory of computing*, pages 419–429, 1991.

[Pel23] Sarah Peluse. Finite field models in arithmetic combinatorics–twenty years on. *arXiv preprint arXiv:2312.08100*, 2023.

[Raz00] Ran Raz. The BNS-Chung criterion for multi-party communication complexity. *Computational Complexity*, 9(2):113–122, 2000.

[RS78] Imre Z Ruzsa and Endre Szemerédi. Triple systems with no six points carrying three triangles. *Combinatorics (Keszthely, 1976), Coll. Math. Soc. J. Bolyai*, 18(939-945):2, 1978.

[RTTV08] Omer Reingold, Luca Trevisan, Madhur Tulsiani, and Salil Vadhan. Dense subsets of pseudorandom sets. In *Proceedings of the 49th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 76–85. IEEE, 2008.

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

[Ser77] Jean-Pierre Serre. *Linear representations of finite groups*, volume 42. Springer, 1977.

[Shk05] Ilya D Shkredov. On a problem of Gowers. In *Doklady Mathematics*, volume 71, pages 46–48. Pleiades Publishing, Ltd., 2005.

[Shk06] Ilya D Shkredov. On a generalization of Szemerédi's theorem. *Proceedings of the London Mathematical Society*, 93(3):723–760, 2006.

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

[Sze75] Endre Szemerédi. *Regular partitions of graphs.* Stanford University, 1975.

[Tao11] Terence Tao. 254B, Notes 3: Quasirandom groups, expansion, and Selberg's 3/16 theorem, December 2011.

[Vio19] Emanuele Viola. Guest column: Non-abelian combinatorics and communication complexity. *ACM SIGACT News*, 50(3):52–74, 2019.

[Zha23] Yufei Zhao. *Graph Theory and Additive Combinatorics: Exploring Structure and Randomness.* Cambridge University Press, 2023.