GAUSSIAN WIDTH BOUNDS WITH APPLICATIONS TO ARITHMETIC PROGRESSIONS IN RANDOM SETTINGS

JOP BRIËT AND SIVAKANTH GOPI

ABSTRACT. Motivated by two problems on arithmetic progressions (APs)—concerning large deviations for AP counts in random sets and random differences in Szemerédi's theorem—we prove upper bounds on the Gaussian width of the image of the n-dimensional Boolean hypercube under a mapping $\psi : \mathbb{R}^n \to \mathbb{R}^k$, where each coordinate is a constant-degree multilinear polynomial with 0/1 coefficients. We show the following applications of our bounds. Let $[\mathbb{Z}/N\mathbb{Z}]_p$ be the random subset of $\mathbb{Z}/N\mathbb{Z}$ containing each element independently with probability p.

- Let X_k be the number of k-term APs in $[\mathbb{Z}/N\mathbb{Z}]_p$. We show that a precise estimate on the large deviation rate $\log \Pr[X_k \geq (1+\delta)\mathbb{E}X_k]$ due to Bhattacharya, Ganguly, Shao and Zhao is valid if $p \geq \omega(N^{-c_k} \log N)$ for $c_k = (6k\lceil (k-1)/2 \rceil)^{-1}$, which slightly improves their bound of $c_k = (6k(k-1))^{-1}$ for $k \geq 5$ (and matching their c_3 and c_4).
- A set $D \subseteq \mathbb{Z}/N\mathbb{Z}$ is *ℓ-intersective* if every dense subset of $\mathbb{Z}/N\mathbb{Z}$ contains a non-trivial $(\ell+1)$-term AP with common difference in D. We show that $[\mathbb{Z}/N\mathbb{Z}]_p$ is ℓ-intersective with probability $1 - o_N(1)$ provided $p \geq \omega(N^{-\beta_\ell} \log N)$ for $\beta_\ell = (\lceil (\ell+1)/2 \rceil)^{-1}$, improving the bound $\beta_\ell = (\ell+1)^{-1}$ due to Frantzikinakis, Lesigne and Wierdl for $\ell \geq 2$ and reproving more directly the same result shown recently by the authors and Dvir.

In addition, we discuss some intriguing connections with special kinds of error correcting codes (locally decodable codes) and the Banach-space notion of type for injective tensor products of ℓ_p-spaces.

1. Introduction

The *Gaussian width* of a point set $T \subseteq \mathbb{R}^k$ measures the average maximum correlation between a standard Gaussian vector $g = N(0, I_k)$ and the points in T,

$$\mathrm{GW}(T) = \mathbb{E}\big[\sup_{t \in T} \langle t, g \rangle\big].$$

The terminology reflects the fact that if T is symmetric around the origin, then its Gaussian width is closely related to its average width in a random direction. Motivated by two applications to arithmetic progressions in random settings discussed below, we bound the Gaussian width of certain sets given by the image of the n-dimensional Boolean hypercube under a polynomial mapping $\psi : \mathbb{R}^n \to \mathbb{R}^k$, where each coordinate ψ_i is a constant-degree multilinear polynomial given by a hypergraph. An s-hypergraph $H = (V, E)$ consists of a vertex set V and a multiset E, also denoted $E(H)$, of subsets of V of size at most s, called the edges. A hypergraph is s-uniform if each edge has size exactly s. The degree of a vertex is the number of edges containing it and the degree of H, denoted $\Delta(H)$, is the maximum degree

J.B. was supported by a VENI grant and the Gravitation-grant NETWORKS-024.002.003 from the Netherlands Organisation for Scientific Research (NWO).

S.G. was supported by NSF CAREER award 1451191 and NSF grant CCF-1523816.

among its vertices. Associate with a hypergraph $H = ([n], E)$, the multilinear polynomial $p_H \in \mathbb{R}[x_1, \dots, x_n]$ given by

$$p_H(x_1, \dots, x_n) = \frac{1}{n} \sum_{e \in E} \prod_{i \in e} x_i. \tag{1}$$

Note that for a subset $A \subseteq [n]$, the value $np_H(1_A)$ counts the number of edges of H which lie completely inside A. Associate with a collection of n-vertex hypergraphs $H_1, \dots, H_k$ the polynomial mapping $\psi_{H_1,\dots,H_k} : \mathbb{R}^n \to \mathbb{R}^k$ given by

$$\psi_{H_1,\dots,H_k}(x) = \begin{pmatrix} p_{H_1}(x) \\ \vdots \\ p_{H_k}(x) \end{pmatrix}. \tag{2}$$

Our main result is then as follows.

Theorem 1.1. *Let k, n, s be positive integers and let $H_1, \dots, H_k$ be s-hypergraphs with vertex set $[n]$. Then,*

$$\mathrm{GW}\left(\psi_{H_1,\dots,H_k}(\{0,1\}^n)\right) \lesssim_s \max_{i \in [k]} \Delta(H_i) \sqrt{kn^{1-\frac{1}{\lceil s/2 \rceil}} \log n}.$$

In the following two subsections we discuss two applications of this result.

1.1. **Large deviations for arithmetic progressions.** Let $H = (V, E)$ be a hypergraph over a finite vertex set V of cardinality N and for $p \in (0, 1)$ denote by V_p the random binomial subset where each element of V appears independently of all others with probability p. Let X be the number of edges in H that are induced by V_p. Important instances of the random variable X include the count of triangles in an Erdős–Rényi random graph and the count of arithmetic progressions of a given length in the random set $[\mathbb{Z}/N\mathbb{Z}]_p$.

The study of the asymptotic behavior of X when $p = p(N)$ is allowed to depend on N and N grows to infinity motivates a large body of research in probabilistic combinatorics. Of particular interest is the problem of determining the probability that X significantly exceeds its expectation $\Pr[X \geq (1 + \delta)\mathbb{E}X]$ for $\delta > 0$, referred to as the *upper tail*. Despite the fact that standard probabilistic methods fail to give satisfactory bounds on the upper tail in general, advances were made recently for special instances, in particular for triangle counts [LZ17] and general subgraph counts [BGLZ17]. For more general hypergraphs, progress was made by Chatterjee and Dembo [CD16] using a novel nonlinear large deviation principle (LDP), which was improved by Eldan [Eld16] shortly after. The LDPs give estimates on the upper tail in terms of a parameter ϕ_p whose value is determined by the solution to a certain variational problem, for a range of values of p depending on H. This splits the problem of estimating the upper tail into two sub-problems: (1) determining for what range of p the estimate in terms of ϕ_p holds true and (2) solving the variational problem to determine the value of ϕ_p. The answer to problem (1) turns out to depend on the Gaussian width of a point set related to H.

This approach was pursued in [CD16] in the context of 3-term arithmetic progressions, for which problem (1) was solved. The case of longer APs was treated by Bhattacharya et al. [BGSZ16], who solved the variational problem (2) and gave bounds for the relevant Gaussian width. Based on this, they showed that for every $k \geq 3$, fixed $\delta > 0$ and p tending to zero sufficiently slowly as $N \to \infty$, the upper tail proability for the count X_k of k-term arithmetic progressions in $[\mathbb{Z}/N\mathbb{Z}]_p$ is given by

$$\Pr[X_k \geq (1+\delta)\mathbb{E}X_k] = p^{(1+o(1))\sqrt{\delta}p^{k/2}N}. \tag{3}$$

The rate at which p is allowed to decay for (3) to hold depends on the Gaussian width of the image of $\{0,1\}^{\mathbb{Z}/N\mathbb{Z}}$ under the gradient $\psi = \nabla p_H$, where H is the hypergraph over $\mathbb{Z}/N\mathbb{Z}$ whose edges are formed by k-term arithmetic progressions. The bounds on the Gaussian width of this set proved in [BGSZ16] imply that (3) holds provided $p \geq N^{-c_k}(\log N)^{\varepsilon_k}$ for

$$c_3 = \frac{1}{18}, \quad c_4 = \frac{1}{48} \quad \text{and} \quad c_k = \frac{1}{6k(k-1)} \quad \text{for } k \geq 5,$$

and absolute constants $\varepsilon_k \in (0,\infty)$ depending only on k. However, the authors conjecture that a probability p slightly larger than $N^{-1/(k-1)}$ suffices. Evidence for this conjecture is given by a result of Warnke [War16] showing that for all $p \geq (\log N/N)^{1/(k-1)}$, the logarithm of the upper tail is given by $\Theta_k(\sqrt{\delta}p^{k/2}N\log p)$, where the asymptotic notation hides constants depending only on k. Notice that (3) improves on this by (almost) determining those constants. The main motivation for finding such precise estimates of the upper tail probability is not so much the problem itself as it is to understand structure of the set $[\mathbb{Z}/N\mathbb{Z}]_p$ conditioned on X_k being much larger than its expectation (see [BGSZ16]). With regard to the constants c_k, Theorem 1.1 implies that for all $k \geq 3$ it suffices to set

$$c_k = \frac{1}{6k\lceil\frac{k-1}{2}\rceil},$$

which slightly improves on the range of p for which (3) was known to hold for $k \geq 5$.

1.2. **Random differences in Szemerédi's Theorem.** In 1975 Szemerédi [Sze75] proved that any subset of the integers of positive upper density contains arbitrarily long arithmetic progressions, answering a famous open question of Erdős and Turán. It is well known that this is equivalent to the assertion that for every positive integer k and any $\alpha \in (0,1)$, there exists an $N_0(k,\alpha) \in \mathbb{N}$ such that if $N \geq N_0(k,\alpha)$ and $A \subseteq [N]$ is a set of size $|A| \geq \alpha N$, then A must contain a non-trivial k-term arithmetic progression. Certain refinements of Szemerédi's theorem concern sets $D \subseteq \mathbb{N}$ such that the theorem still holds true when the arithmetic progressions are required to have common difference from D. Such sets are usually referred to as intersective sets in number theory, or recurrent sets in ergodic theory. More precisely, a set $D \subseteq \mathbb{N}$ is ℓ-intersective (or ℓ-recurrent) if any set $A \subseteq \mathbb{N}$ of positive upper density has an $(\ell+1)$-term arithmetic progression with common difference in D. Szemerédi's theorem then states that $\mathbb{N}$ is ℓ-intersective for every $\ell \in \mathbb{N}$, but much smaller intersective sets exist. For example, for any $t \in \mathbb{N}$, the set $\{1^t, 2^t, 3^t, \dots\}$ is ℓ-intersective for every ℓ, which is a special case of more general results of Sárközy [Sár78a] when $\ell = 1$ and of Bergelson and Leibman [BL96] for all $\ell \geq 1$. The shifted primes $\{p-1 : p \text{ is prime}\}$ and $\{p+1 : p \text{ is prime}\}$ are also ℓ-intersective for every $\ell \in \mathbb{N}$, shown by Sárközy [Sár78b] when $\ell = 1$ and in a more general setting by Wooley and Ziegler [WZ12] for all $\ell \geq 1$.

It is natural to ask at what density, random sets become ℓ-intersective. To simplify the discussion, we will look at the analogous question in $\mathbb{Z}/N\mathbb{Z}$.

Definition 1.2. Let ℓ be a positive integer and $\alpha \in (0,1]$. A subset $D \subseteq \mathbb{Z}/N\mathbb{Z}$ is (ℓ,α)-intersective if any subset $A \subseteq \mathbb{Z}/N\mathbb{Z}$ of size $|A| \geq \alpha N$ must contain a non-trivial $(\ell+1)$-term arithmetic progression with common difference in D.

It was proved independently by Frantzikinakis et al. [FLW12] and Christ [Chr11] that for $\beta_\ell = \frac{1}{2^{\ell-1}}$ and $p \geq \omega(N^{-\beta_\ell} \log N)$, the random set $[\mathbb{Z}/N\mathbb{Z}]_p$ is (ℓ,α)-intersective with probability $1 - o_N(1)$, provided $N \geq N_1(\ell,\alpha)$. This was improved for all $\ell \geq 2$ in [FLW16], where it was shown that the same result holds with $\beta_\ell = \frac{1}{\ell+1}$, though it was conjectured there that $\beta_\ell = 1$ suffices for all $\ell \geq 1$. Based on Theorem 1.1 we obtain the following result, which improves on the latter bounds.

Theorem 1.3. *For every $\ell \in \mathbb{N}, \alpha \in (0,1)$ there exists an $N_1(\ell,\alpha) \in \mathbb{N}$ such that the following holds. Let $N > N_1(\ell,\alpha)$ be an integer and let*

$$\beta_\ell = \frac{1}{\lceil \frac{\ell+1}{2} \rceil} \quad \textit{and} \quad p \geq \omega(N^{-\beta_\ell} \log N).$$

Then, with probability $1 - o_N(1)$, the set $[\mathbb{Z}/N\mathbb{Z}]_p$ is (ℓ,α)-intersective.

1.3. **Locally decodable codes.** There is a close connection between the Gaussian widths considered in Theorem 1.1 and special error-correcting codes known as *locally decodable codes* (LDCs). A map $C : \{0,1\}^k \to \{0,1\}^n$ is a q-query LDC if for every $i \in [k]$ and $x \in \{0,1\}^k$, the value x_i can be retrieved by reading at most q coordinates of the codeword $C(x)$, even if the codeword is corrupted in a not too large (but possibly constant) fraction of coordinates. A main open problem is to determine the optimal trade-off between n and k when q is a fixed constant. Currently this problem is settled only in the cases $q = 1, 2$ [KT00, KW04, GKST06] and remains wide open for the case $q = 3$. We refer to the extensive survey [Yek12] for more information on this problem. The connection with Gaussian width was established by the authors and Dvir in [BDG17], where we showed that there is a q-query LDC from $\{0,1\}^{\Omega(k)}$ to $\{0,1\}^{O(n)}$ if and only if there are q-matchings $H_1, \dots, H_k$ on $[n]$ of size $\Omega(n)$ such that the set $\psi_{H_1,\dots,H_k}(\{0,1\}^n)$ has Gaussian width $\Omega(k)$. It was observed there that the best-known lower bounds on the length $n = n(k)$ of q-query LDCs—proved using techniques from quantum information theory [KW04]—imply a slightly different but equivalent version of Theorem 1.3 (see Section 5). The proof of Theorem 1.1 is based on ideas from [KW04], but uses a 1974 random matrix inequality of Tomczak–Jaegermann instead of quantum information theory.[1]

1.4. **Gaussian width bounds from type constants.** We observe that the Gaussian width in Theorem 1.1 can be bounded in terms of type constants of certain Banach spaces. Unfortunately, we do not have good enough bounds on the type constants of the required spaces to improve Theorem 1.1. But we hope that this connection will motivate progress on understanding these spaces.

A Banach space X is said to have (Rademacher) type $p > 0$ if there exists a constant $T < \infty$ such that for every k and $x_1, \dots, x_k \in X$,

$$\mathbb{E}_\varepsilon \left\| \sum_{i=1}^k \varepsilon_i x_i \right\|_X^p \leq T^p \sum_{i=1}^k \|x_i\|_X^p \,, \tag{4}$$

[1]Not surprisingly, the LDC lower bounds of [KW04] are also implied by Theorem 1.1.

where the expectation is over a uniformly random $\varepsilon = (\varepsilon_1, \dots, \varepsilon_k) \in \{-1, 1\}^k$. The smallest T for which (4) holds is referred to as the type-p constant of X, denoted $T_p(X)$. Type, and its dual notion cotype, play an important role in Banach space theory as they are tightly linked to local geometric properties (we refer to [LT79] and [Mau03] for extensive surveys). Some fundamental facts are as follows. It follows from the triangle inequality that every Banach space has type 1 and from the Khintchine inequality that no Banach space has type $p > 2$. The parallelogram law implies that Hilbert spaces have type 2. An easy but important fact is that ℓ_1 fails to have type $p > 1$. Indeed, a famous result of Maurey and Pisier [MP73] asserts that a Banach space fails to have type $p > 1$ if and only if it contains ℓ_1 uniformly. Finite-dimensional Banach spaces have type-p for all $p \in [1, 2]$. But of importance to Theorem 1.1 are the actual type constants $T_p(X)$ of a certain family of finite-dimensional Banach spaces. Let $r_1, \dots, r_s \geq 1$ be such that $\sum_{i=1}^s \frac{1}{r_i} = 1$ and let $\mathcal{L}^n_{r_1,\dots,r_s}$ be the space of s-linear forms on $\mathbb{R}^n \times \cdots \times \mathbb{R}^n$ (s times) endowed with the norm

$$\|\Lambda\| = \sup\Big\{ \frac{|\Lambda(x_1, \dots, x_s)|}{\|x_1\|_{\ell_{r_1}} \cdots \|x_s\|_{\ell_{r_s}}} : x_1, \dots, x_s \in \mathbb{R}^n \setminus \{0\} \Big\}.$$

This space is also known as the injective tensor product of $\ell^n_{r_1'}, \dots, \ell^n_{r_s'}$ for $\frac{1}{r_i} + \frac{1}{r_i'} = 1$ and as such plays an important role in the theory of tensor products of Banach spaces [Rya02]. The relevance of the type constants of this space to Theorem 1.1 is captured by the following lemma, proved in Section 6.

Lemma 1.4. *Let k, n, s be positive integers and let $H_1, \dots, H_k$ be s-hypergraphs with vertex set $[n]$. Then for any $r_1, \dots, r_s \geq 1$ such that $\sum_{i=1}^s \frac{1}{r_i} = 1$ and any $p \in [1, 2]$,*

$$\mathrm{GW}\big(\psi_{H_1,\dots,H_k}(\{0,1\}^n)\big) \lesssim_s T_p(\mathcal{L}^n_{r_1,\dots,r_s})\, k^{1/p} \max_{i \in [k]} \Delta(H_i).$$

Observe that the space $\mathcal{L}^n_{2,2}$ may be identified with the space of $n \times n$ matrices endowed with the spectral norm (or operator norm). A key ingredient in the proof of Theorem 1.1, Theorem 2.2 below, easily implies that the type-2 constant of this space is of order $O(\sqrt{\log n})$. A well-known lower bound of the same order follows for instance from the connection between Gaussian width and LDCs and a basic construction of a 2-query LDC known as the Hadamard code. More generally, lower bounds on the type constants of $\mathcal{L}^n_{r_1,\dots,r_s}$ are implied by s-query LDCs [BNR12, Bri16].

Acknowledgements. We thank Sean Prendiville, Fernando Xuancheng Shao and Yufei Zhao for helpful discussions. This work was in part carried out while the authors were visiting the Simons Institute during the Pseudorandomness program of Spring 2017 and we thank the institute and organizers for their hospitality.

2. Proof of Theorem 1.1

In this section we prove Theorem 1.1. We begin by giving a high-level overview of the ideas. The proof is based on a classic random matrix inequality of Tomczak-Jaegermann which bounds the expected operator norm of a sum of matrices weighted by independent standard normal random variables (Theorem 2.2 below). On its own, this inequality easily implies the result for graphs. To treat the general case, we first reduce to the case of $2r$-uniform matchings for $r = \lceil s/2 \rceil$, at a cost of a factor $O(r \max_i \Delta(H_i))$ in the number of vertices; a *matching* is a hypergraph where no two edges intersect. Then we reduce to the case of graphs (unless $r = 1$) and apply the random matrix inequality. This involves constructing graphs G_i on approximately $n^{n^{1-1/r}}$ vertices, with the property that for each $x \in \{-1,1\}^n$ it holds that $p_{H_i}(x) = cp_{G_i}(x^{\otimes n^{1-1/r}})$ for some constant c depending only on r (Lemma 2.3). (Switching from Boolean vectors to sign vectors can only make the Gaussian width larger.) To illustrate how these graphs are constructed, we consider a 4-matching H on n vertices and let $N = n^{\sqrt{n}}$. It follows from the Birthday Paradox and symmetry that the number the strings in $[n]^{\sqrt{n}}$ containing at least two elements of a given edge $e \in E(H)$ is $\Omega(N/n)$. We let G be the graph with vertex set $[n]^{\sqrt{n}}$ with the edges formed by the strings (u,v) that "cover" some edge in H and "complement" each other, meaning: there are indices $i,j \in [\sqrt{n}]$ such that $\{u_i, u_j, v_i, v_j\} \in E(H)$ and $u_\ell = v_\ell$ for all $\ell \notin \{i,j\}$. The m-fold tensor product of a vector $x \in \mathbb{R}^n$ is given by $x^{\otimes m} = (\prod_{i=1}^m x_{u_i})_{u\in[n]^m}$. If $\{u,v\}$ is an edge in G and $x \in \{-1,1\}^n$, it then follows that $(x^{\otimes\sqrt{n}})_u (x^{\otimes\sqrt{n}})_v = x_{u_i}x_{u_j}x_{v_i}x_{v_j}$. It can then be observed that $p_G(x^{\otimes\sqrt{n}})$, modulo the relations $x_1^2 = 1, \dots, x_n^2 = 1$, is a linear combination of the monomials appearing in $p_H(x)$. A more careful analysis shows that the two evaluations are in fact related by a constant factor.

To make the above precise, we first collect some basic facts about hypergraphs. The *edge chromatic number* of a hypergraph H, denoted by $\chi_E(H)$, is the minimum number of colors needed to color the edges of H such that no two edges which intersect have the same color. Note that $\chi_E(H)$ equals the smallest number of matchings into which $E(H)$ can be partitioned. For small values of s, the parameters $\chi_H(E)$ and $\Delta(H)$ are closely related.

Lemma 2.1. *Let H be an s-hypergraph. Then,*

$$\Delta(H) \le \chi_E(H) \le s(\Delta(H) - 1) + 1.$$

Proof: Clearly $\chi_E(H) \ge \Delta(H)$ since edges containing a maximum degree vertex should get different colors. To prove the upper bound, form a graph G whose vertices are $E(H)$, and add edges between intersecting hypergraph edges. Then $\chi_E(H)$ is equal to the vertex chromatic number of the graph G, which, by Brooks' Theorem, is at most $\Delta(G) + 1$. Since an edge in H can intersect at most $s(\Delta(H) - 1)$ other edges, $\Delta(G) \le s(\Delta(H) - 1)$. □

The proof of Theorem 1.1 uses a non-commutative Khintchine inequality, which is a special case of a result of Tomczak-Jaegermann [TJ74, Theorem 3.1]. Let $\langle \cdot, \cdot \rangle$ be the standard inner product on $\mathbb{R}^N$ and denote by B_2^N the Euclidean unit ball in $\mathbb{R}^N$. Given a matrix $A \in \mathbb{R}^{N\times N}$, its operator norm (or spectral norm) is given by $\|A\| = \sup\{|\langle x, Ay\rangle| : x, y \in B_2^N\}$.

Theorem 2.2 (Tomczak-Jaegermann)**.** *There exists an absolute constant $C \in (0,\infty)$ such that the following holds. Let $A_1, \dots, A_k \in \mathbb{R}^{N\times N}$ be a collection of matrices and let $g_1, \dots, g_k$ be independent Gaussian random variables with mean zero and variance 1. Then,*

$$\mathbb{E}\Big[\Big\|\sum_{i=1}^k g_i A_i\Big\|\Big] \le C\sqrt{\log N}\Big(\sum_{i=1}^k \|A_i\|^2\Big)^{1/2}.$$

To apply Theorem 2.2 we use the following matrix lemma, proved in the next section.

Lemma 2.3. *For every $r \in \mathbb{N}$ there exist a $C_r, c_r \in (0,\infty)$ and $n_0(r) \in \mathbb{N}$ such that the following holds. Let $n \ge n_0(r)$, $m = C_r n^{1-1/r}$ and $N = n^m$. Let $H = ([n], E)$ be a $2r$-uniform* *hypergraph and let p_H be the polynomial as in (1). Then, there exists a matrix $A \in \mathbb{R}^{N\times N}$ such that $\|A\| \lesssim_r \Delta(H)$ and for every $x \in \{-1,1\}^n$,*

$$p_H(x) = \frac{1}{c_r N}\langle x^{\otimes m}, Ax^{\otimes m}\rangle.$$

Moreover, A is the adjacency matrix of a graph (with possible parallel edges).

Proof of Theorem 1.1: Assume $n \geq n_0(r)$ as in Lemma 2.3. Let $K = \max_{i\in[k]} \Delta(H_i)$ and $r = \lceil s/2 \rceil$. We start by reducing to the setting where $H_1, \ldots, H_k$ are $2r$-uniform of degree at most K. Given an s-hypergraph H over $[n]$ with degree at most K, we first add new vertices to edges with less than $2r$ vertices to make all the edges have size $2r$. By grouping the edges into n sets of size at most K and using the same new vertices for all edges in the same group, we can obtain a new hypergraph H' on $n' = O(rn)$ vertices which is $2r$-uniform and whose degree is still at most K. In terms of the polynomials, we are homogenizing the polynomial $np_H(x)$ with new variables to a get a new multilinear polynomial $n'p_{H'}(x, x')$ of degree $2r$ where x' are the new variables added. Clearly, $np_H(x) = n'p_{H'}(x, x')$ when the x' variables are set to 1. So,

$$\mathrm{GW}\left(\psi_{H_1,\ldots,H_k}(\{0,1\}^n)\right) \lesssim_r \mathrm{GW}\left(\psi_{H'_1,\ldots,H'_k}(\{0,1\}^{n'})\right).$$

Since our claimed bound on the Gaussian width is $O(n)$, the extra vertices will result in at most an extra factor $O(r)$. It thus suffices to prove the theorem for the case where $H_1, \ldots, H_k$ are $2r$-uniform. Also observe that since the polynomials p_{H_i} are multilinear, the Gaussian width is bounded from above by replacing binary vectors with sign vectors

$$\mathrm{GW}\left(\psi_{H_1,\ldots,H_k}(\{0,1\}^n)\right) \leq \mathbb{E}\max\Big\{\sum_{i=1}^k g_i p_{H_i}(x) : x \in \{-1,1\}^n\Big\}.$$

Let $m = C_r n^{1-1/r}$ and $N = n^m$ and for each $i \in [k]$, let $A_i \in \mathbb{R}^{N\times N}$ be a matrix for pH_i as in Lemma 2.3. Then, for every $x \in \{-1,1\}^n$,

$$\begin{aligned}
\sum_{i=1}^k g_i p_{H_i}(x) &= \frac{1}{c_r N}\sum_{i=1}^n g_i \langle x^{\otimes m}, A_i x^{\otimes m}\rangle \\
&= c_r^{-1}\Big\langle \frac{x^{\otimes m}}{\sqrt{N}}, \Big(\sum_{i=1}^k g_i A_i\Big)\frac{x^{\otimes m}}{\sqrt{N}}\Big\rangle \\
&\leq c_r^{-1}\Big\|\sum_{i=1}^k g_i A_i\Big\|.
\end{aligned}$$

Hence, by Theorem 2.2,

$$\begin{aligned}
\mathrm{GW}\left(\psi_{H_1,\ldots,H_k}(\{0,1\}^n)\right) &\leq c_r^{-1}\mathbb{E}\Big[\Big\|\sum_{i=1}^k g_i A_i\Big\|\Big] \\
&\leq c_r^{-1}\sqrt{\log N}\Big(\sum_{i=1}^k \|A_i\|^2\Big)^{1/2} \\
&\lesssim_r K\sqrt{kn^{1-1/r}\log n},
\end{aligned}$$

where in the last line we used that $\|A_i\| \le O_r(K)$ for each $i \in [k]$. $\square$

3. Matrix lemma

Here we prove Lemma 2.3. Let $\mathcal{M} \subseteq \binom{[n]}{2r}$ be a maximal family of disjoint $2r$-sets of $[n]$. Let $t = 200 \cdot 4^r$. Given a string $x \in \{-1,1\}^n$ write its m-fold tensor product as

$$x^{\otimes m} = \Big(\prod_{i=1}^{m} x_{f(i)}\Big)_{f:[m]\to[n]}.$$

Given a mapping $f : [m] \to [n]$ and set $S \in \mathcal{M}$, let

$$\mu_S(f) = \sum_{T \in \binom{S}{r}} \prod_{i \in T} |f^{-1}(i)|.$$

Note that this is a count of the r-subsets $I \subseteq [m]$ such that $|S \cap f(I)| = r$. Denote

$$\phi(f) = \sum_{S \in \mathcal{M}} \mu_S(f).$$

For $\ell \in \mathbb{N}$, say that f is *ℓ-good* if $1 \le \phi(f) \le \ell$. Say that $g : [m] \to [n]$ *complements* f if it satisfies the following two criteria:

(1) There exists exactly one $I \in \binom{[m]}{r}$ such that $f(I) \cup g(I) \in \mathcal{M}$.
(2) For all $i \in [m] \setminus I$, we have $g(i) = f(i)$.

If g complements f then clearly the converse also holds. Say that the complementary pair (f,g) *covers* $S \in \mathcal{M}$ if $f(I) \cup g(I) = S$. Observe that if (f,g) covers S, then for every $x \in \{-1,1\}^m$, we have

$$(x^{\otimes m})_f (x^{\otimes m})_g = \prod_{i=1}^{m} x_{f(i)} x_{g(i)} = \prod_{j \in S} x_j. \tag{5}$$

Define the set of ordered pairs

$$\mathcal{P} = \big\{(f,g) : \ f \text{ is } t\text{-good and } g \text{ complements } f\big\}. \tag{6}$$

Proposition 3.1. *Let $\mathcal{P}$ be as in (6). Then, for every $S \in \mathcal{M}$, the number of pairs $(f,g) \in \mathcal{P}$ that cover S equals $|\mathcal{P}|/|\mathcal{M}|$.*

Proof: Fix distinct sets $S, T \in \mathcal{M}$ and let $\pi \in S_n$ be a permutation such that $\pi(S) = T, \pi(T) = S$ and $\pi(i) = i$ for all $i \notin S \cup T$. Let $\mathcal{P}_S$ be the set of pairs $(f,g) \in \mathcal{P}$ which cover S and define $\mathcal{P}_T$ similarly. We claim that the map $\psi : (f,g) \mapsto (\pi \circ f, \pi \circ g)$ is an injective map from $\mathcal{P}_S$ to $\mathcal{P}_T$. It follows that T is covered by at least as many pairs from $\mathcal{P}$ as S is. Similarly, interchanging S and T, the converse also holds. To prove the claim, note that if (f,g) covers S, then $(\pi \circ f, \pi \circ g)$ covers T. Moreover, $\phi(\pi \circ f) = \phi(f)$ because π maps edges of the matching $\mathcal{M}$ to edges of $\mathcal{M}$. Thus $\psi(\mathcal{P}_S) \subset \mathcal{P}_T$. Finally ψ is injective because if $\pi \circ f = \pi \circ f'$ for some $f, f' : [m] \to [n]$, then $f = f'$. Hence $\mathcal{P}$ covers all $S \in \mathcal{M}$ equally. $\square$

Proposition 3.2. *For every $(f,g) \in \mathcal{P}$, we have that g is t^2-good.*

Proof: Let $S \in \mathcal{M}$ and $(f,g) \in \mathcal{P}$ be such that (f,g) covers S. Consider the histograms $F, G : [n] \to \{0,1,\ldots,m\}$ given by $F(i) = |f^{-1}(i)|$ and $G(i) = |g^{-1}(i)|$ for each $i \in [n]$. Then F and G differ only in S. In particular, there is an r-set $T \subseteq S$ such that $G(i) = F(i)+1$ for each $i \in T$ and $G(i) = F(i) - 1$ for each $i \in S \setminus T$. Hence,

$$\begin{aligned}
\mu_S(g) &= \sum_{T \in \binom{S}{r}} \prod_{i \in T} G(i) \\
&\leq \sum_{T \in \binom{S}{r}} \prod_{i \in T} \big(F(i) + 1\big) \\
&\leq \sum_{T \in \binom{S}{r}} \Big(1 + 2^r \prod_{i \in T} F(i)\Big) \\
&\leq 4^r + 2^r \mu_S(f).
\end{aligned}$$

For all other $S' \in \mathcal{M}$, we have $\mu_{S'}(g) = \mu_{S'}(f)$. Moreover, f must be t-good for (f,g) to belong to $\mathcal{P}$. It follows that

$$\phi(g) = \sum_{S' \in \mathcal{M}} \mu_{S'}(g) \leq 4^r + 2^r \sum_{S' \in \mathcal{M}} \mu_{S'}(f) = 4^r + 2^r \phi(f) \leq t^2,$$

where in the last line we used the choice of $t = 200 \cdot 4^r$. □

Lemma 3.3 (Generalized birthday paradox)**.** *For every $r \in \mathbb{N}$ there exists a $C_r \in (0,\infty)$ and an $n_0(r) \in \mathbb{N}$ such that the following holds. Let h be a uniformly distributed random variable over the set of maps from $[m]$ to $[n]$. Then, provided $n \geq n_0(r)$ and $m = C_r n^{1-1/r}$,*

$$\Pr\big[h \text{ is } t\text{-good}\big] \geq \frac{1}{2}.$$

We postpone the proof of Lemma 3.3 to Section 3.1.

Corollary 3.4. *Let $\mathcal{P}$ be as in (6) and let $A : [n]^m \times [n]^m \to \{0,1\}$ be its incidence matrix, that is $A(f,g) = 1 \iff (f,g) \in \mathcal{P}$. Then, $|\mathcal{P}| \geq \Omega(N)$ and every row and every column of A has at most $t^2(r!)$ ones.*

Proof: The first claim follows from Lemma 3.3 and the fact that $|\mathcal{P}|$ is at least the number of t-good mappings. If h is l-good, then there are at most $l(r!)$ mappings from $[m] \to [n]$ that complement h. Hence, every row of A has at most $t(r!)$ ones and by Proposition 3.2, every column of A has at most $t^2(r!)$ ones. □

Proof of Lemma 2.3: By Lemma 2.1, the hypergraph can be decomposed into $\chi_E(H) \leq 2rK$ matchings, which we denote by $\mathcal{F}_1, \ldots, \mathcal{F}_{\chi_E(H)}$. Complete each $\mathcal{F}_i$ to a maximal family $\mathcal{M}_i$ of disjoint $2r$-subsets of $[n]$ in some arbitrary way. For each $\mathcal{M}_i$, let $\mathcal{P}_i$ be as in (6) and let $A_i : [n]^m \times [n]^m \to \{0,1\}^n$ be its incidence matrix. Set to zero all the entries of A_i that correspond to a pair (f,g) covering a set in $\mathcal{M}_i \setminus \mathcal{F}_i$. Let $B = A_1 + \cdots + A_{\chi_E(H)}$ and $A = (B + B^{\mathsf{T}})$. It follows from (5) and Proposition 3.1 that for every $x \in \{-1,1\}^n$, we have

$$\Big\langle x^{\otimes m}, \sum_{i=1}^{\chi_E(H)} (A_i + A_i^{\mathsf{T}}) x^{\otimes m} \Big\rangle = 2 \sum_{i=1}^{\chi_E(H)} \frac{|\mathcal{P}_i|}{|\mathcal{M}_i|} \sum_{S \in \mathcal{F}_i} \prod_{j \in S} x_i. \tag{7}$$

Since all $\mathcal{M}_i$ are maximal, they have the same size, as do the $\mathcal{P}_i$. Hence, by Corollary 3.4, there exists a constant $c_r \in (0,1]$ such that the right-hand side of (7) equals $(2c_rN/n)p(x)$. Let G be the graph with adjacency matrix A, allowing for parallel edges. Then G has degree at most $2Kt^2(r!)$ and, it follows from Lemma 2.1 that G can be partitioned into $O_r(K)$ matchings. Since the adjacency matrix of a matching has unit norm, we get that $\|A\| \leq O_r(K)$. $\square$

3.1. **Proof of the generalized birthday paradox.** For the proof of Lemma 3.3, we use a standard Poisson approximation result for "balls and bins" problems [MU05, Theorem 5.10]. A discrete Poisson random variable Y with expectation μ is nonnegative, integer valued, and has probability density function

$$\Pr[Y = \ell] = \frac{e^{-\mu}\mu^\ell}{\ell!}, \qquad \forall \ell = 0, 1, 2, \ldots \tag{8}$$

Proposition 3.5. *If X, Y are independent Poisson random variables with expectations μ_X, μ_Y, respectively, then $X + Y$ is a Poisson random variable with expectation $\mu_X + \mu_Y$.*

Lemma 3.6. *Let h be a uniformly distributed map from $[m]$ to $[n]$. For each $i \in [n]$, let $X_i = |h^{-1}(i)|$ and let $\mathbf{X} = (X_i)_{i\in[n]}$. Let $\mathbf{Y} = (Y_i)_{i\in[n]}$ be a vector of independent Poisson random variables with expectation m/n. Then, for any nonnegative function $\Phi : (\mathbb{N} \cup \{0\})^n \to \mathbb{R}_+$ such that $\mathbb{E}[\Phi(\mathbf{X})]$ decreases or increases monotonically with m, we have*

$$\mathbb{E}[\Phi(\mathbf{X})] \leq 2\mathbb{E}[\Phi(\mathbf{Y})].$$

Proof of Lemma 3.3: Let $C_r > 0$ be a parameter depending only on r to be set later. Let $\mu = C_r m/n = C_r n^{-1/r}$ and assume that $n \geq n_0(r) := 4(C_r r)^r$. For h a random map as in Lemma 3.6, we begin by lower bounding the probability of the event that $\phi(h) \geq 1$. Recall that this occurs if there exists an $S \in \mathcal{M}$ and an r-subset $T \in \binom{S}{r}$ such that $T \subseteq \mathrm{im}(h)$. Let $\mathbf{X}$ be as in Lemma 3.6. Let $\psi : (\mathbb{N} \cup \{0\})^n \to \{0,1\}$ be the function

$$\psi(x) = \prod_{S\in\mathcal{M}} \prod_{T\in\binom{S}{r}} \Big(1 - \prod_{i\in T} 1_{\geq 1}(x_i)\Big).$$

Then $\psi(\mathbf{X}) = 1$ if $\phi(h) = 0$ and $\psi(\mathbf{X})$ decreases monotonically with m. Hence, for $\mathbf{Y}$ a Poisson random vector as in Lemma 3.6, we have

$$\begin{aligned}
\Pr[\phi(h) = 0] &= \mathbb{E}[\psi(\mathbf{X})] \\
&\leq 2\mathbb{E}[\psi(\mathbf{Y})] \\
&= 2\prod_{S\in\mathcal{M}} \mathbb{E}\Big[\prod_{T\in\binom{S}{r}} \Big(1 - \prod_{i\in T} 1_{\geq 1}(Y_i)\Big)\Big],
\end{aligned} \tag{9}$$

where in the last line we used the fact that since the sets $S \in \mathcal{M}$ are disjoint, the random variables

$$\prod_{T\in\binom{S}{r}} \Big(1 - \prod_{i\in T} 1_{\geq 1}(Y_i)\Big)$$

are independent. The random variables $1_{\geq 1}(Y_i)$, $i \in S$, are independent Bernoullis that are zero with probability $e^{-\mu}$. The expectation in (9) equals the probability that these random variables form a string of Hamming weight strictly less than r. Using that $n \geq 4(C_r r)^r$ and the fact that $1 - x \leq \exp(-x) \leq 1 - x + x^2/2$ when $x > 0$, this probability is at most

$$1 - \Pr[\forall i \in T \ 1_{\geq 1}(Y_i) = 1] = 1 - (1 - e^{-\mu})^r \leq 1 - (\mu(1 - \mu/2))^r \leq 1 - \frac{C_r^r}{en} \leq \exp\Big(-\frac{C_r^r}{en}\Big)$$

where $T \subset S$ is some fixed subset of size r. Hence, since $\mathcal{M}$ is maximal, the above and (9) give

$$\Pr[\phi(h) = 0] \leq 2\exp\left(-\frac{C_r^r|\mathcal{M}|}{en}\right) \leq 2\exp\left(-\frac{C_r^r\lfloor n/r\rfloor}{en}\right) \leq 2\exp\left(-\frac{C_r^r}{2er}\right). \tag{10}$$

Set $C_r = (6er)^{1/r}$, then the above right-hand side is at most $1/4$. Next, we upper bound the probability that $\phi(h) \geq t = 200 \cdot 4^r$. Define $\chi : (\mathbb{N} \cup \{0\})^n \to \mathbb{R}_+$ by

$$\chi(x) = \sum_{S \in \mathcal{M}} \sum_{T \in \binom{S}{r}} \prod_{i \in T} x_i.$$

Then, $\phi(h) = \chi(\mathbf{X})$. Moreover, $\mathbb{E}[\chi(\mathbf{X})]$ increases monotonically with m. It thus follows from Lemma 3.6 that

$$\begin{aligned}
\mathbb{E}[\phi(h)] &\leq 2\mathbb{E}[\chi(\mathbf{Y})] \\
&= 2\sum_{S \in \mathcal{M}} \sum_{T \in \binom{S}{r}} \prod_{i \in T} \mathbb{E}[Y_i] \\
&\leq 2|\mathcal{M}|\binom{2r}{r}\left(\frac{m}{n}\right)^r \\
&\leq 2 \cdot \frac{n}{r} \cdot 4^r \cdot (6er)n^{-1} \leq 50 \cdot 4^r.
\end{aligned}$$

where in the second line we used the fact that the Y_i are independent. By Markov's inequality, $\Pr[\phi(h) > 200 \cdot 4^r] \leq \frac{1}{4}$. With (10), we get that h is t-good with probability at least $1/2$. $\square$

4. Arithmetic progressions in random sets

Below we state a special case of Eldan's LDP [Eld16], similar to how it is stated in [BGSZ16]. Consider a multilinear polynomial $F \in \mathbb{R}[x_1, \dots, x_N]$ with zero constant term. The discrete Lipschitz constant of F is given by

$$\mathrm{Lip}(F) = \max\big\{\|(\nabla F)(y)\|_{\ell_\infty} : y \in \{0,1\}^N\big\},$$

where $\nabla F : \mathbb{R}^N \to \mathbb{R}^N$ is the gradient of F. For $p, q \in [0,1]$, define

$$I_p(q) = q\log\tfrac{q}{p} + (1-q)\log\tfrac{1-q}{1-p}.$$

For a vector $q \in [0,1]^N$, let $Y_q = (Y_1, \dots, Y_N)$ be a random vector of independent random variables $Y_i \sim \mathrm{Bernoulli}(q_i)$. Define $\phi_p : \mathbb{R}_+ \to \mathbb{R}_+$ by

$$\phi_p(t) = \inf_{q \in [0,1]^N} \Big\{\sum_{i=1}^N I_p(q_i) : \mathbb{E}F(Y_q) \geq tN\Big\}. \tag{11}$$

Theorem 4.1 (Eldan). *Let $X = (X_1, \dots, X_N)$ be a vector of independent Bernoulli(p) random variables and let $F \in \mathbb{R}[x_1, \dots, x_N]$ be a multilinear form with zero constant term. Let t, ε be real numbers such that $N^{-1}\phi_p(t-\varepsilon) > \varepsilon > 0$. Then,*

$$\log \Pr[F(X) \geq tN] \leq -\Big(1 - \tfrac{6L(\log N)^{1/6}}{\varepsilon N^{1/3}}\Big)\phi_p(t-\varepsilon),$$

where

$$L = \Big(\big(2 + \tfrac{1}{\varepsilon\sqrt{N}}\big)\operatorname{Lip}(F) + |\log(p(1-p))|\Big)^{2/3}\Big(\operatorname{GW}\big((\nabla F)(\{0,1\}^N)\big) + \tfrac{1}{\varepsilon}\operatorname{Lip}(F)^2\Big)^{1/3}. \tag{12}$$

Moreover, if $2\operatorname{Lip}(F)^2/(\varepsilon^2 N) \leq 1$, then

$$\log \Pr[F(X) \geq (t-\varepsilon)N] \geq -\Big(1 + \tfrac{2\operatorname{Lip}(F)^2}{\varepsilon^2 N}\Big)\phi_p(t) - \log 10.$$

Theorem 1.1 can be applied to get an upper bound on the parameter L given by (12) when F is a polynomial p_H given in terms of a hypergraph H with the property that only few edges are incident on any two vertices. For example, if the edges are the (unordered) $(r+1)$-term arithmetic progressions in $\mathbb{Z}/N\mathbb{Z}$ with non-zero common difference, then any pair of distinct vertices $i, j \in \mathbb{Z}/N\mathbb{Z}$ appears in at most r^2 edges. Indeed, if $i < j$ and both belong to an $(r+1)$-term AP, then there is a step count $\ell \in [r]$ and common difference $d \in \mathbb{Z}/N\mathbb{Z} \setminus \{0\}$ such that $j = i + \ell d$. This leaves at most r possibilities for d and r possibilities for the position of i in the AP.

Proposition 4.2. *Let N, K, r be positive integers. Let $H = ([N], E)$ be a $(r+1)$-hypergraph such that at most K edges are incident on any given pair of vertices. Then, $\operatorname{Lip}(p_H) \leq K$ and*

$$\operatorname{GW}\big((\nabla p_H)(\{0,1\}^N)\big) \lesssim_r K N^{1-\frac{1}{2\lceil r/2\rceil}}\sqrt{\log N}.$$

Proof: For each $i \in [N]$ let $H_i = ([N], E_i)$ be the r-hypergraph with edge set

$$E_i = \{e \setminus \{i\} : e \in E(H) \text{ and } i \in e\}.$$

Then each H_i has degree at most K and $p_{H_i} = \nabla(p_H)_i$. Since H_i has at most KN edges, for every $y \in \{0,1\}^N$ and $i \in [N]$, we have $p_{H_i}(y) \leq K$. This implies the first claim. The second claim follows from Theorem 1.1, since $\nabla p_H = \psi_{H_1,\dots,H_N}$. □

By applying Eldan's LDP with the function

$$\Lambda_k(x) = \frac{1}{N}\sum_{a,b \in \mathbb{Z}/N\mathbb{Z}, b \neq 0} x_a x_{a+b} x_{a+2b} \cdots x_{a+(k-1)b}.$$

and explicitly solving the variational problem (11) for this specific function, Bhattacharya et al. [BGSZ16] obtained the following general upper tail estimate on the count of k-term arithmetic progressions in $[\mathbb{Z}/N\mathbb{Z}]_p$.

Theorem 4.3 (Bhattacharya–Ganguly–Shao–Zhao). *Let $k \geq 3$ be an integer and let σ, τ be positive real numbers such that*

$$\operatorname{GW}\big((\nabla \Lambda_k)(\{0,1\}^N)\big) \lesssim N^{1-\sigma}(\log N)^{\tau}.$$

Let $p \in (0,1)$ be bounded away from 1 and let $\delta > 0$ be such that $\delta = O(1)$ and

$$\min\{\delta p^k, \delta^2 p\} \gtrsim N^{-\sigma/3}(\log N)^{1+\tau/3}.$$

Then,

$$\log \Pr[\Lambda_k([\mathbb{Z}/N\mathbb{Z}]_p) \geq (1+\delta)\mathbb{E}\Lambda_k([\mathbb{Z}/N\mathbb{Z}]_p)] = -\big(1+o(1)\big)\,\phi_p\big((1+o(1))\delta\big). \tag{13}$$

Moreover, provided $\delta p^k N^2 \to \infty$, we have

$$\phi_p(\delta) \asymp N\,\min\{\sqrt{\delta}p^{k/2}, \delta^2 p\}.$$

By Proposition 4.2, we may set $\sigma = 1/(2\lceil (k-1)/2\rceil)$ and $\tau = 1/2$ in Theorem 4.3.

Corollary 4.4. *The upper tail estimate* (13) *holds when*

$$\min\{\delta p^k, \delta^2 p\} \gtrsim N^{-\frac{1}{6\lceil (k-1)/2\rceil}} (\log N)^{1+1/6}.$$

5. Random differences in Szemerédi's Theorem

In this section, we will prove Theorem 1.3. We will first consider a slightly different random model where we form a random multiset D_k of size k by repeatedly sampling a uniformly random element from $\mathbb{Z}/N\mathbb{Z}$ for k times. We will need the following equivalent formulation of Szemerédi's Theorem due to Varnavides [Var59] (see [Tao07, Theorem 4.8] for this exact formulation).

Proposition 5.1. *For every $\ell \in \mathbb{N}, \alpha \in (0,1]$ there exists $N_1(\ell,\alpha), \epsilon(\ell,\alpha)$ such that for every $N \geq N_1(\ell,\alpha)$, the following holds. Every subset $A \subseteq \mathbb{Z}/N\mathbb{Z}$ of size at least αN contains an $\epsilon(\ell,\alpha)$-fraction of all $\ell+1$ term arithmetic progressions in $\mathbb{Z}/N\mathbb{Z}$, that is,*

$$\mathbb{E}_{x\in\mathbb{Z}/N\mathbb{Z}, y\in\mathbb{Z}/N\mathbb{Z}\setminus\{0\}}[1_A(x)1_A(x+y)\dots 1_A(x+\ell y)] \geq \epsilon(\ell,\alpha).$$

Proposition 5.2. *For all $\ell \in \mathbb{N}, \alpha \in (0,1]$ there exists $N_1(\ell,\alpha) \in \mathbb{N}$ such that for every $N > N_1(\ell,\alpha)$ the following holds. Let $k \geq \omega(N^{1-1/\lceil(\ell+1)/2\rceil}\log N)$ and let D be a random multiset of size k obtained by sampling k times independently and uniformly at random from $\mathbb{Z}/N\mathbb{Z}\setminus\{0\}$. Then, with probability $1-o_N(1)$, every subset $A \subseteq \mathbb{Z}/N\mathbb{Z}$ of size at least αN contains a non-trivial arithmetic progression of length $\ell+1$ with common difference in D.*

Proof: Denote $\Gamma = \mathbb{Z}/N\mathbb{Z}$. We will arrive at a contradiction assuming that the statement is false. Let $N_1(\ell,\alpha)$ and $\epsilon(\ell,\alpha)$ be as in Proposition 5.1. Suppose that with a constant probability, there is a subset $A \subseteq N$ of size at least αN with no non-trivial $\ell+1$ arithmetic progression whose common difference lies in D. Then,

$$\Pr_D\left[\inf_{A:|A|\geq\alpha N} \mathbb{E}_{y\in D}\mathbb{E}_{x\in\Gamma}[1_A(x)1_A(x+y)\dots 1_A(x+\ell y)] = 0\right] = \Omega(1).$$

By Proposition 5.1, for every $A \subset N$ of size at least αN,

$$\mathbb{E}_{y\in\Gamma\setminus\{0\}}\mathbb{E}_{x\in\Gamma}[1_A(x)1_A(x+y)\dots 1_A(x+\ell y)] \geq \epsilon.$$

For $f : \Gamma \to \mathbb{R}$, define $\phi_y(f) = \mathbb{E}_{x\in\Gamma}[f(x)f(x+y)\dots f(x+\ell y)]$ which is a degree $\ell+1$ polynomial over the variables $(f(x))_{x\in\Gamma}$. Let $\sigma_1,\dots,\sigma_k$ be independent uniformly distributed $\{-1,1\}$-valued random variables and let D' be an independent copy of D. Combining both the observations and using a standard symmetrization trick, we get:

$$\begin{aligned}
\epsilon &\lesssim \mathbb{E}_D\Big[\sup_{A:|A|\geq\alpha N}\left|\mathbb{E}_{y\in D}\phi_y(1_A) - \mathbb{E}_{y\in\Gamma\setminus\{0\}}\phi_y(1_A)\right|\Big]\\
&= \mathbb{E}_D\Big[\sup_{A:|A|\geq\alpha N}\left|\mathbb{E}_{y\in D}\phi_y(1_A) - \mathbb{E}_{D'}\mathbb{E}_{y\in D'}\phi_y(1_A)\right|\Big]\\
&\leq \mathbb{E}_{D,D'}\Big[\sup_{A:|A|\geq\alpha N}\left|\mathbb{E}_{y\in D}\phi_y(1_A) - \mathbb{E}_{y\in D'}\phi_y(1_A)\right|\Big]\\
&= \mathbb{E}_{y_1,\dots,y_k\in\Gamma\setminus\{0\};\ y_1',\dots,y_k'\in\Gamma\setminus\{0\}}\mathbb{E}_\sigma\Big[\sup_{A:|A|\geq\alpha N}\left|\frac{1}{k}\sum_{i=1}^k\sigma_i\left(\phi_{y_i}(1_A)-\phi_{y_i'}(1_A)\right)\right|\Big]\\
&\leq 2\mathbb{E}_{y_1,\dots,y_k\in\Gamma\setminus\{0\}}\mathbb{E}_\sigma\Big[\sup_{A:|A|\geq\alpha N}\left|\frac{1}{k}\sum_{i=1}^k\sigma_i\phi_{y_i}(1_A)\right|\Big].
\end{aligned}$$

Let us fix $y_1,\dots,y_k\in\Gamma\setminus\{0\}$. Each ϕ_{y_i} can be written as $\phi_{y_i} = p_{H_i}$ where H_i is the hypergraph on Γ whose edges are given by $(\ell+1)$ term arithmetic progressions with common difference y_i. The maximum degree of H_i is $O(\ell)$. This is because each such AP $(x+ty_i)_{0\leq t\leq\ell}$ intersects another AP $(x'+t'y_i)_{0\leq t'\leq\ell}$ iff $x-x'=(t'-t)y_i$; so there are only $O(\ell)$ such x' for a given x. Therefore, by applying Theorem 1.1, we get

$$\mathrm{GW}(\psi_{H_1,\dots,H_k}(\{0,1\}^n)) \lesssim_\ell \sqrt{kN^{1-1/\lceil(\ell+1)/2\rceil}\log N}.$$

Let $g_1,\dots,g_k$ be independent $N(0,1)$ random variables. Then we can bound

$$\begin{aligned}
\mathbb{E}_\sigma\left[\sup_{A:|A|\geq\alpha N}\left|\frac{1}{k}\sum_{i=1}^k\sigma_i\phi_{y_i}(1_A)\right|\right] &\lesssim \frac{1}{k}\mathbb{E}_g\left[\sup_A\left|\sum_{i=1}^k g_i\phi_{y_i}(1_A)\right|\right]\\
&= \frac{1}{k}\mathbb{E}_g\left[\sup_A\left|\sum_{i=1}^k g_i p_{H_i}(1_A)\right|\right]\\
&= \frac{1}{k}\,\mathrm{GW}(\psi_{H_1,\dots,H_k}(\{0,1\}^n))\\
&\leq \frac{1}{k}\sqrt{kN^{1-1/\lceil(\ell+1)/2\rceil}\log N}.
\end{aligned}$$

Thus we get $k\lesssim_\ell N^{1-1/\lceil(\ell+1)/2\rceil}\log N$ which is a contradiction. $\square$

We will the need following simple fact that conditioning on a high probability event will not change the probability of any event by much.

Lemma 5.3. *Let A,E be some events in some probability space. If $\Pr[E]\geq 1-\varepsilon$ then $|\Pr[A|E]-\Pr[A]|\leq 2\varepsilon/(1-\varepsilon)$.*

Proof:

$$
\begin{aligned}
|\Pr[A|E] - \Pr[A]| &= \left| \frac{\Pr[A \cap E]}{\Pr[E]} - \Pr[A] \right| \\
&= \left| \frac{1}{\Pr[E]} \left(\Pr[A] + \Pr[E] - \Pr[A \cup E] \right) - \Pr[A] \right| \\
&\le \left| \Pr[A] \left(\frac{1}{\Pr[E]} - 1 \right) \right| + \left| 1 - \frac{\Pr[A \cup E]}{\Pr[E]} \right| \\
&\le \frac{2\varepsilon}{1-\varepsilon}
\end{aligned}
$$

□

Proof of Theorem 1.3: Let D_k be a random subset of $\mathbb{Z}/N\mathbb{Z} \setminus \{0\}$ of size at most k, formed by sampling a uniformly random element from $\mathbb{Z}/N\mathbb{Z}$ for k times. Let $D_p = [\mathbb{Z}/N\mathbb{Z} \setminus \{0\}]_p$ be a random subset of $\mathbb{Z}/N\mathbb{Z} \setminus \{0\}$ formed by including each element with probability p independently. We claim that if D_k is ℓ-intersective with probability $1 - o_N(1)$, then D_p will also be ℓ-intersective with probability $1 - o_N(1)$ when $p = 2k/N$ and $k = \omega_N(1)$.

Let $p = 2k/N$ and $k = \omega_N(1)$. Let E be the event that D_p has size at least k. By the Chernoff bound,

$$
1 - \Pr[E] \le \exp\left(-\mathrm{D}_{\mathrm{KL}} \left(\frac{p}{2} || p \right) N \right) \le \exp(-\Omega(pN)) = o_N(1)
$$

where D_{KL} is the Kullback-Leibler divergence. By Lemma 5.3, conditioning on E changes the probability of D_p being ℓ-intersective by $o_N(1)$. Conditioned on E, the probability that D_p is ℓ-intersective is at least the probability that D_k is ℓ-intersective. Indeed, both D_p and D_k, after conditioning on a given size reduce to the uniform distribution over all subsets of that size. Proposition 5.2 thus implies D_p is ℓ-intersective when $p = \omega(N^{-1/\lceil (\ell+1)/2 \rceil} \log N)$. □

6. Proof of Lemma 1.4

In this section we give a proof Lemma 1.4. As explained in the proof of Theorem 1.1, it suffices to prove the statement for s-uniform hypergraphs $H_1, \ldots, H_k$. The coordinates of $\psi_{H_1,\ldots,H_k}$ are given by $p_{H_1}, \ldots, p_{H_k}$ defined as in (1). Each $p_{H_i}(x)$ is a degree-s homogeneous polynomial in $x = (x_1, \ldots, x_n)$. Let Λ_{H_i} be an s-multilinear form such that $p_{H_i}(x) = \Lambda_{H_i}(x, x, \ldots, x)$. Let $g = (g_1, \ldots, g_k)$ be vector of independent standard Gaussians and

$\varepsilon = (\varepsilon_1, \ldots, \varepsilon_k)$ be uniformly random in $\{-1,1\}^k$. Then

$$\begin{aligned}
\mathrm{GW}\left(\psi_{H_1,\ldots,H_k}(\{0,1\}^n)\right) &= \mathbb{E}_g \sup_{x\in\{0,1\}^n} \left|\sum_{i=1}^k g_i p_{H_i}(x)\right| \\
&= \mathbb{E}_g \sup_{x\in\{0,1\}^n} \left|\sum_{i=1}^k g_i \Lambda_{H_i}(x,\ldots,x)\right| \\
&\leq \mathbb{E}_g n^{\sum_{i=1}^k 1/r_i} \left\|\sum_{i=1}^k g_i \Lambda_{H_i}\right\| \\
&\leq n\mathbb{E}_g\mathbb{E}_\varepsilon \left\|\sum_{i=1}^k \varepsilon_i g_i \Lambda_{H_i}\right\|,
\end{aligned}$$

where in the last line we used that each g_i is symmetrically distributed, that is, g_i and $-g_i$ have the same distribution. By Jensen's inequality, the above expectation over ε is at most

$$\left(\mathbb{E}_\varepsilon \left\|\sum_{i=1}^k \varepsilon_i g_i \Lambda_{H_i}\right\|^p\right)^{1/p} \leq T_p(\mathcal{L}^n_{r_1,\ldots,r_s}) \left(\sum_{i-1}^k \|g_i\Lambda_{H_i}\|^p\right)^{1/p},$$

where the inequality follows from the definition of the type-p constant of $\mathcal{L}^n_{r_1,\ldots,r_s}$. Hence,

$$\begin{aligned}
\mathrm{GW}\left(\psi_{H_1,\ldots,H_k}(\{0,1\}^n)\right) &\leq n\mathbb{E}_g\ T_p(\mathcal{L}^n_{r_1,\ldots,r_s}) \left(\sum_{i=1}^k \|g_i\Lambda_{H_i}\|^p\right)^{1/p} \\
&\leq nT_p(\mathcal{L}^n_{r_1,\ldots,r_s})\mathbb{E}_g \|g\|_{\ell_p} \max_i \|\Lambda_{H_i}\| \\
&\leq nT_p(\mathcal{L}^n_{r_1,\ldots,r_s})\, k^{1/p} \max_i \|\Lambda_{H_i}\|,
\end{aligned}$$

where we used the fact that $\mathbb{E}_g \|g\|_{\ell_p} \leq (\sum_{i=1}^k \mathbb{E}_{g_i}|g_i|^p)^{1/p} \leq k^{1/p}(\mathbb{E}_{g_1}|g_1|^2)^{1/2} = k^{1/p}$. If H_i is a matching hypergraph, using Hölder's inequality, it is easy to see that $\|\Lambda_{H_i}\| \leq 1/n$. If not, by Lemma 2.1, we can decompose H_i into $s\Delta(H_i)$ matchings and use triangle inequality to conclude that $\|\Lambda_{H_i}\| \leq s\Delta(H_i)/n$ which gives the desired bound. □

References

[BDG17] Jop Briët, Zeev Dvir, and Sivakanth Gopi. Outlaw distributions and locally decodable codes. In *Proceedings of the 8th Annual Innovations in Theoretical Computer Science conference (ITCS 2017)*, 2017. To appear. arXiv preprint: 1609.06355.

[BGLZ17] Bhaswar Bhattacharya, Shirshendu Ganguly, Eyal Lubetzky, and Yufei Zhao. Upper tails and independence polynomials in random graphs. *Advances in Mathematics*, 319:313–347, 2017.

[BGSZ16] Bhaswar Bhattacharya, Shirshendu Ganguly, Xuancheng Shao, and Yufei Zhao. Upper tails for arithmetic progressions in a random set. *arXiv preprint: 1605.02994*, 2016.

[BL96] Vitaly Bergelson and Alexander Leibman. Polynomial extensions of van der Waerden's and Szemerédis theorems. *Journal of the American Mathematical Society*, 9(3):725–753, 1996.

[BNR12] Jop Briët, Assaf Naor, and Oded Regev. Locally decodable codes and the failure of cotype for projective tensor products. *Electronic Research Announcements in Mathematical Sciences (ERA-MS)*, 19:120–130, 2012.

[Bri16] Jop Briët. On embeddings of ℓ_1^k from locally decodable codes. *arXiv preprint: arXiv:1611.06385*, 2016.

[CD16] Sourav Chatterjee and Amir Dembo. Nonlinear large deviations. *Advances in Mathematics*, 299:396–450, 2016.

[Chr11] Michael Christ. On random multilinear operator inequalities. *arXiv preprint: 1108.5655*, 2011.

[Eld16] Ronen Eldan. Gaussian-width gradient complexity, reverse log-Sobolev inequalities and nonlinear large deviations. *arXiv preprint: 1612.04346*, 2016.

[FLW12] Nikos Frantzikinakis, Emmanuel Lesigne, and Mate Wierdl. Random sequences and pointwise convergence of multiple ergodic averages. *Indiana University Mathematics Journal*, pages 585–617, 2012.

[FLW16] Nikos Frantzikinakis, Emmanuel Lesigne, and Mate Wierdl. Random differences in Szemerédi's theorem and related results. *Journal d'Analyse Mathématique*, 130(1):91–133, 2016.

[GKST06] Oded Goldreich, Howard Karloff, Leonard J Schulman, and Luca Trevisan. Lower bounds for linear locally decodable codes and private information retrieval. *Computational Complexity*, 15(3):263–296, 2006. Preliminary version appeared in CCC'02.

[KT00] Jonathan Katz and Luca Trevisan. On the efficiency of local decoding procedures for error-correcting codes. In *Proceedings of the 32nd annual ACM symposium on Theory of computing (STOC 2000)*, pages 80–86. ACM Press, 2000.

[KW04] Iordanis Kerenidis and Ronald de Wolf. Exponential lower bound for 2-query locally decodable codes via a quantum argument. *J. of Computer and System Sciences*, 69:395–420, 2004. Preliminary version appeared in STOC'03.

[LT79] Joram Lindenstrauss and Lior Tzafriri. *Classical Banach spaces. II*, volume 97 of *Ergebnisse der Mathematik und ihrer Grenzgebiete [Results in Mathematics and Related Areas]*. Springer-Verlag, Berlin-New York, 1979. Function spaces.

[LZ17] Eyal Lubetzky and Yufei Zhao. On the variational problem for upper tails in sparse random graphs. *Random Structures & Algorithms*, 50(3):420–436, 2017.

[Mau03] Bernard Maurey. Type, cotype and K-convexity. In *Handbook of the geometry of Banach spaces, Vol. 2*, pages 1299–1332. North-Holland, Amsterdam, 2003.

[MP73] Bernard Maurey and Gilles Pisier. Caractérisation d'une classe d'espaces de Banach par des propriétés de séries aléatoires vectorielles. *C. R. Acad. Sci. Paris Sér A*, 277:687—690, 1973.

[MU05] Michael Mitzenmacher and Eli Upfal. *Probability and computing: Randomized algorithms and probabilistic analysis*. Cambridge University Press, 2005.

[Rya02] Raymond A. Ryan. *Introduction to tensor products of Banach spaces*. Springer Monographs in Mathematics. Springer-Verlag London Ltd., London, 2002.

[Sár78a] András Sárközy. On difference sets of sequences of integers. I. *Acta Mathematica Hungarica*, 31(1-2):125–149, 1978.

[Sár78b] András Sárközy. On difference sets of sequences of integers. III. *Acta Mathematica Hungarica*, 31(3-4):355–386, 1978.

[Sze75] Endre Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arith*, 27(299-345):21, 1975.

[Tao07] Terence Tao. The ergodic and combinatorial approaches to Szemerédi's theorem. In *CRM Proc. Lecture Notes*, volume 43, pages 145–193, 2007.

[TJ74] Nicole Tomczak-Jaegermann. The moduli of smoothness and convexity and the Rademacher averages of trace classes S_p $(1 \leq p < \infty)$. *Studia Math.*, 50:163–182, 1974.

[Var59] Panayiotis Varnavides. On certain sets of positive density. *Journal of the London Mathematical Society*, 1(3):358–360, 1959.

[War16] Lutz Warnke. Upper tails for arithmetic progressions in random subsets. *arXiv preprint: 1612.08559*, 2016.

[WZ12] Trevor Wooley and Tamar Ziegler. Multiple recurrence and convergence along the primes. *American Journal of Mathematics*, 134(6):1705–1732, 2012.

[Yek12] Sergey Yekhanin. Locally decodable codes. *Foundations and Trends in Theoretical Computer Science*, 6(3):139–255, 2012.

CWI, Science Park 123, 1098 XG Amsterdam, The Netherlands
E-mail address: j.briet@cwi.nl

Department of Computer Science, Princeton University, Princeton, NJ 08540, USA
E-mail address: sgopi@cs.princeton.edu