GAUSSIAN WIDTH BOUNDS WITH APPLICATIONS TO ARITHMETIC PROGRESSIONS IN RANDOM SETTINGS

JOP BRIËT AND SIVAKANTH GOPI

Abstract. Motivated by a problem on random differences in Szemerédi's theorem and another problem on large deviations for arithmetic progressions in random sets, we prove upper bounds on the Gaussian width of special point sets in $\mathbb{R}^k$. The point sets are formed by 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.

- A set $D \subseteq \mathbb{Z}/N\mathbb{Z}$ is *ℓ-intersective* if any dense subset of $\mathbb{Z}/N\mathbb{Z}$ contains a proper $(\ell+1)$-term arithmetic progression with common difference in D. Our main result implies that $[\mathbb{Z}/N\mathbb{Z}]_p$ is ℓ-intersective with probability $1-o(1)$ provided $p \geq \omega(N^{-\beta_\ell}\log N)$ for $\beta_\ell = (\lceil(\ell+1)/2\rceil)^{-1}$. This gives a polynomial improvement for all $\ell \geq 2$ of a previous bound due to Frantzikinakis, Lesigne and Wierdl, and reproves more directly the same improvement shown recently by the authors and Dvir (here we avoid the theories of locally decodable codes and quantum information).
- Let X_k be the number of k-term arithmetic progressions in $[\mathbb{Z}/N\mathbb{Z}]_p$ and consider the large deviation rate $\rho_k(\delta) = \log \Pr[X_k \geq (1+\delta)\mathbb{E}X_k]$. We give quadratic improvements of the best-known range of p for which a highly precise estimate of $\rho_k(\delta)$ due to Bhattacharya, Ganguly, Shao and Zhao is valid for all odd $k \geq 5$. In particular, the estimate holds if $p \geq \omega(N^{-c_k}\log N)$ for $c_k = (6k\lceil(k-1)/2\rceil)^{-1}$.

We also discuss connections with the above-mentioned 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 expected maximum correlation between T and a standard Gaussian vector $g = N(0, I_k)$, and is given by

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

The terminology reflects the fact that the Gaussian width of a set is proportional to $\sqrt{k}$ times its average width in a random direction. While this quantity plays a central role in high-dimensional probability, it is notoriously hard to estimate in general; see for instance [Tal14] for an extensive discussion of this problem.

Our main result gives upper bounds on the Gaussian width of sets that appear naturally in the context of probabilistic combinatorics. The relevant sets are given by the image of the n-dimensional Boolean hypercube under a certain polynomial mapping $\psi : \mathbb{R}^n \to \mathbb{R}^k$. In particular, we focus on the case where each coordinate $\psi_i : \mathbb{R}^n \to \mathbb{R}$ is a multilinear

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.

polynomial with 0-1 coefficients. Say that a polynomial has *multiplicity* t if each of its variables has a nonzero exponent in at most t monomials in its support.[1]

Theorem 1.1. *Let $\psi : \mathbb{R}^n \to \mathbb{R}^k$ be a polynomial mapping such that each coordinate is multilinear, has 0-1 coefficients, and has degree at most d and multiplicity t. Then,*

$$\mathrm{GW}\left(\psi(\{0,1\}^n)\right) \lesssim_d nt\sqrt{kn^{1-\frac{1}{\lceil d/2\rceil}}\log n}.$$

The factor nt can be seen as a natural scaling due to the fact that each coordinate ψ_i maps the Boolean hypercube into $[0, nt]$ (which follows from a handshaking lemma). In the special case where ψ is linear, $\psi(x) = (\langle c_1, x\rangle, \dots, \langle c_k, x\rangle)$, for some $c_1, \dots, c_k \in \{0,1\}^N$, the set $\psi(\{0,1\}^n)$ is easily seen to be contained in the set $T = \{(\langle c_i, y\rangle)_{i=1}^k : \|y\|_{\ell_\infty} \le 1\}$. The Gaussian width of the former set is thus at most that of the latter, which in turn is at most

$$\mathbb{E}\Big[\Big\|\sum_{i=1}^k g_i c_i\Big\|_{\ell_1}\Big] \lesssim n\sqrt{k},$$

as the sum is an n-dimensional Gaussian vector whose coordinates have variance at most k. Perhaps surprisingly, Theorem 1.1 shows that if ψ is quadratic and has constant multiplicity, then the Gaussian width is at most a factor $\sqrt{\log n}$ larger than the above upper bound. This turns out to be an easy consequence of a 1974 random matrix inequality due to Tomczak–Jeagermann [TJ74], which also forms the basis for our proof of the higher-degree cases. The proof of Theorem 1.1 (given in Section 2) proceeds in two steps: first we reduce to the case of homogeneous mappings of even degree, and then we reduce to the quadratic case. The first step is the reason for the ceiling in $\lceil d/2\rceil$ appearing in the exponent and it would be interesting to know if one can remove this ceiling (i.e., does the result hold with the exponent $1 - 2/d$?). Finally, a close inspection of the proof of Theorem 1.1 shows that it also holds for polynomials with non-negative integer coefficients, for a suitable change of the definition of multiplicity. In the following four subsections we discuss two applications of this result and links with error correcting codes and the Banach space notion of type.

1.1. 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 $N \ge N_0(k,\alpha)$ and $A \subseteq \mathbb{Z}/N\mathbb{Z}$ is a set of size $|A| \ge \alpha N$, then A must contain a proper k-term arithmetic progression. Certain refinements of Szemerédi's theorem concern sets $D \subseteq \mathbb{N}$ for which 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 $\ell \ge 1$. The shifted primes $\{p-1 : p \text{ is prime}\}$ and $\{p+1 : p \text{ is prime}\}$

[1] Here and below, $\lesssim, \gtrsim$ denote upper and lower bounds up-to absolute constants and $\lesssim_\varepsilon, \gtrsim_\varepsilon$ denote upper and lower bounds up-to constants depending on a parameter ε.

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 $|A| \geq \alpha N$ contains a proper $(\ell+1)$-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 $1 - o(1)$, 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 $\ell \in \mathbb{N}$ and $\alpha \in (0,1)$, there exists an $N_1(\ell, \alpha) \in \mathbb{N}$ such that the following holds. Let $N \geq N_1(\ell, \alpha)$ 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 $1 - o(1)$, the set $[\mathbb{Z}/N\mathbb{Z}]_p$ is (ℓ, α)-intersective.

1.2. **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 precise estimates on the upper tail that are given in terms of a parameter ϕ_p whose value is determined by the solution to a certain variational problem. The range of values of p for which these estimates are actually valid depends on the underlying hypergraph 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] to estimate the upper tail of the number of 3-term arithmetic progressions in $[\mathbb{Z}/N\mathbb{Z}]_p$, for which the authors solved problem (1). The case of longer APs, asking for the upper tail probability of the count X_k of k-term arithmetic progressions in $[\mathbb{Z}/N\mathbb{Z}]_p$, was recently treated by Bhattacharya et al. [BGSZ18]. They solved the variational problem (2) for N prime and gave bounds for the relevant Gaussian width towards solving problem (1). Based on this, they showed that if $k \geq 3$ and $\delta > 0$ are fixed and p tends to zero sufficiently slowly as $N \to \infty$ along the primes, then

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

Similar results were shown for the analogous problem over $\{1, \ldots, N\}$ (in which case N no longer needs to be prime), but we shall focus on the problem in $\mathbb{Z}/N\mathbb{Z}$ for ease of exposition. The rate at which p is allowed to decay for (1) to hold turns out to depend on Gaussian widths of the form featuring in Theorem 1.1. The bounds proved in [BGSZ18] imply that (1) 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 for all k. Some support 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 (also referred to as the large deviation rate) of the k-AP count in $\{1, \ldots, N\}_p$ 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 (1) is more accurate than this result in that it (almost) determines those constants, though currently for a more narrow range of p.[2] Using Theorem 1.1, we widen the range of p for which (1) can be shown to hold for all $k \geq 5$.

Theorem 1.4. *For every integer $k \geq 3$ and*

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

the estimate (1) *holds true, provided $p \geq N^{-c_k}(\log N)$ and N is prime.*

1.3. Locally decodable codes.

There is a close connection between the Gaussian widths considered in Theorem 1.1 and special error-correcting codes called *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 smallest possible codeword length n as a function of the message length 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. A connection with Gaussian width was established by the authors and Dvir in [BDG17], where we show that q-query LDCs from $\{0,1\}^{\Omega(k)}$ to $\{0,1\}^{O(n)}$ are equivalent to mappings $\psi : \mathbb{R}^n \to \mathbb{R}^k$ whose coordinates are degree-q, multiplicity-1 polynomials with 0-1 coefficients that are supported by $\Omega(n)$ monomials, and such that the set $\psi(\{0,1\}^n)$ 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

[2] 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 [BGSZ18]).

different but equivalent version of Theorem 1.3 (see Section 5). The proof of Theorem 1.1 is based on ideas from [KW04], but does not use quantum information theory.[3]

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 $x_1, \dots, x_k \in X$,

$$\mathbb{E}_\varepsilon \left\| \sum_{i=1}^k \varepsilon_i x_i \right\|_X^p \leq T^p \sum_{i=1}^k \|x_i\|_X^p , \tag{2}$$

where the expectation is over a uniformly random $\varepsilon = (\varepsilon_1, \dots, \varepsilon_k) \in \{-1, 1\}^k$. The smallest T for which (2) 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 $p \in [1, 2]$.

Of importance to Theorem 1.1 are the actual type constants $T_p(X)$ of a certain family of finite-dimensional Banach spaces. Let $r_1, \dots, r_d \geq 1$ be such that $\sum_{i=1}^d \frac{1}{r_i} = 1$ and let $\mathcal{L}^n_{r_1,\dots,r_d}$ be the space of d-linear forms on $\mathbb{R}^n \times \cdots \times \mathbb{R}^n$ (d times) endowed with the norm

$$\|\Lambda\| = \sup \left\{ \frac{|\Lambda(x_1, \dots, x_d)|}{\|x_1\|_{\ell_{r_1}} \cdots \|x_d\|_{\ell_{r_d}}} : x_1, \dots, x_d \in \mathbb{R}^n \setminus \{0\} \right\}.$$

This space is also known as the injective tensor product of $\ell^n_{s_1}, \dots, \ell^n_{s_d}$ for $r_i^{-1} + s_i^{-1} = 1$ 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 7.

Lemma 1.5. *Let $\psi : \mathbb{R}^n \to \mathbb{R}^k$ be a polynomial mapping such that each coordinate is multilinear and has 0-1 coefficients, degree at most d and multiplicity t. Then for any $r_1, \dots, r_d \geq 1$ such that $\sum_{i=1}^d \frac{1}{r_i} = 1$ and any $p \in [1, 2]$,*

$$\mathrm{GW}\big(\psi(\{0,1\}^n)\big) \lesssim_d nt\, T_p(\mathcal{L}^n_{r_1,\dots,r_d})\, k^{1/p}.$$

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 2.1 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

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

Hadamard code. More generally, lower bounds on the type constants of $\mathcal{L}^n_{r_1,\dots,r_d}$ are implied by d-query LDCs [BNR12, Bri16].

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 main tool we use is the following random matrix inequality, which is a special case of a non-commutative version of the Khintchine inequality due to 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 $\|A\| = \sup\{|\langle Ax, y\rangle| : x, y \in B_2^N\}$.

Theorem 2.1 (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,*

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

This result already suffices to prove Theorem 1.1 when the coordinate mappings ψ_i are quadratic forms, in which case there exist matrices $A_i \in \{0,1\}^{n\times n}$ such that $\psi_i(x) = \langle A_i x, x\rangle$. The assumption that each ψ_i has multiplicity t implies that each row and column of A_i has at most t ones. This in turn implies that $\|A_i\| \le t$ by a Birkhoff-von Neumann-type theorem. Since each $x \in \{0,1\}^n$ has Euclidean norm at most $\sqrt{n}$, we get

$$\mathrm{GW}\big(\psi(\{0,1\}^n)\big) = \mathbb{E}\Big[\max_{x\in\{0,1\}^n} \sum_{i=1}^k g_i\langle A_i x, x\rangle\Big] = \mathbb{E}\Big[\max_{x\in\{0,1\}^n} \Big\langle \Big(\sum_{i=1}^k g_i A_i\Big)x, x\Big\rangle\Big] \le n\mathbb{E}\Big[\Big\|\sum_{i=1}^k g_i A_i\Big\|\Big].$$

By Theorem 2.1, the above is at most $Ctn\sqrt{k\log n}$.

The general case is proved via a reduction to the above quadratic case and consists of two steps. In the first step, we reduce to the case where each coordinate ψ_i is a homogeneous polynomial of degree $2\lceil d/2\rceil$. This is done in a straightforward way by adding at most dn variables in such a way so as to preserve the multiplicity. The second step consists of a reduction to the quadratic case. For this, it will be convenient to consider the hypergraphs associated with the monomial support of the coordinate mappings ψ_i.

Recall that an d-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 d, called the edges. A hypergraph is d-uniform if each edge has size exactly d. The degree of a vertex is the number of edges containing it and the degree of H, denoted $\Delta(H)$, is the maximum degree among its vertices. A *matching* is a hypergraph where no two edges intersect. Associate with a hypergraph $H = ([n], E)$, the multilinear polynomial $p_H \in \mathbb{R}[x_1, \dots, x_n]$ given by

$$p_H(x_1,\dots,x_n) = \sum_{e\in E}\prod_{i\in e} x_i. \tag{3}$$

The multiplicity of p_H is then exactly the degree $\Delta(H)$. Clearly the coordinate mappings ψ_i of the form featuring in Theorem 1.1 can be written as p_H for some d-hypergraph H of degree at most t. The reduction to the quadratic case is based on the following key lemma, in which for $x \in \mathbb{R}^n$ and $m \in \mathbb{N}$, the the mth tensor power x is defined as $x^{\otimes m} = (\prod_{i=1}^m x_{u_i})_{u\in[n]^m}$.

Lemma 2.2 (Matrix lemma)**.** *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 $n \geq n_0(r)$, $m = C_r n^{1-1/r}$ and $N = n^m$. Let $H = ([n], E)$ be a $2r$-uniform hypergraph and let p_H be the polynomial as in* (3)*. Then, there exists a matrix $A \in \mathbb{R}^{N\times N}$ such that $\|A\| \lesssim_r \Delta(H)$ and for every $x \in \{-1,1\}^n$,*

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

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

With this lemma in hand, the proof of Theorem 1.1 is straightforward (see below). The idea behind Lemma 2.2 is to use decompositions into matchings and a generalization of the Birthday Paradox that says that for any n-vertex $2r$-matching, a random subset of $C_r n^{1-1/r}$ vertices contains r vertices of any fixed edge with probability c_r/n. To illustrate how this is used in the $r = 2$ case, let H be a 4-matching, let $m = C_2\sqrt{n}$ and $N = n^m$. It follows from the generalized Birthday Paradox that there are $c_2 N/n$ strings in $[n]^m$ containing at least two elements of a given edge. Now let G be the graph with vertex set $[n]^m$ whose edges are the pairs $\{u, v\}$ that *cover* some edge in H and *complement* each other, meaning: there are indices $i, j \in [m]$ 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 main observation is that for every edge $\{u, v\} \in E(G)$ that covers an edge $e \in E(H)$ and every $x \in \{-1,1\}^n$, we have

$$(x^{\otimes m})_u(x^{\otimes m})_v = \prod_{\ell=1}^{m} x_{u_\ell}x_{v_\ell} = x_{u_i}x_{u_j}x_{v_i}x_{v_j} = \prod_{w\in e} x_w.$$

It follows that, modulo the relations $x_1^2 = 1, \dots, x_n^2 = 1$, we have $p_G(x^{\otimes m}) = (c_2 N/n)p_H(x)$. The lemma would now follow by letting A be the appropriately scaled adjacency matrix of G, were it not for the issue that G could have very high degree, which would result in A having a large operator norm. To deal with this, we instead consider a pruned version of G in which we keep only edges that do not cover too many edges of H.

We now give the formal proof of Theorem 1.1. The following simple proposition is used for the first step, in which we homogenize the polynomials. Given two hypergraphs H, H', say that H' *majorizes* H if $V(H) \subseteq V(H')$ and if for each edge $e \in E(H)$, there is a unique edge $e' \in E(H')$ such that $e \subseteq e'$.

Proposition 2.3. *For any n-vertex d-hypergraph H, there is a d-uniform hypergraph H' on dn vertices that majorizes H and satisfies $\Delta(H') = \Delta(H)$.*

Proof: Let $t = \Delta(H)$. It follows from the handshaking lemma that $|E(H)| \leq tn$. Partition $E(H) = \{E_1, \dots, E_n\}$ into n pairwise disjoint sets of size at most t each. Add to $V(H)$ pairwise disjoint sets $W_1, \dots, W_n$ of $d-1$ new vertices each. For each $i \in [n]$, complete each edge $e \in E_i$ to a set of size d by adding vertices from W_i and let H' be the hypergraph thus obtained. Observe that we have not increased the degree of the vertices in $V(H)$. Since each E_i has size at most t, the new vertices in W_i also have degree at most t and therefore, $\Delta(H') = t$. It is trivial to verify that H' satisfies the other desired properties. □

Proof of Theorem 1.1: Let $r = \lceil d/2\rceil$ and for each $i \in [k]$, let H_i be the d-hypergraph of degree t such that $\psi_i = p_{H_i}$, with p_{H_i} as in (3). Assume that $n \geq n_0(r)$ for $n_0(r)$ as in Lemma 2.2. We start by reducing to the setting where each H_i is $2r$-uniform and of degree at most t. To this end, let $H_i' = ([n] \cup [(2r-1)n], E_i')$ be a $2r$-uniform hypergraph that majorizes H_i as in Proposition 2.3, which exists since any d-hypergraph is a $2r$-hypergraph. Then, for each $e \in E(H_i)$, there is a unique set $f(e) \subseteq [(2r-1)n]$ such that $e \cup f(e) \in E(H_i')$. It follows that

$$p_{H_i}(x) = \sum_{e \in E(H_i)} \prod_{i \in e} x_i = \sum_{e \in E(H_i)} \prod_{i \in e} x_i \prod_{j \in f(e)} 1 = p_{H_i'}((x, \mathbf{1})),$$

where $\mathbf{1} \in \mathbb{R}^{(2r-1)n}$ is the all-ones vector. Hence, if we let $\psi' : \mathbb{R}^{2rn} \to \mathbb{R}^k$ be the polynomial map whose coefficients are given by $p_{H_i'}$, then

$$\mathrm{GW}\left(\psi(\{0,1\}^n)\right) \leq \mathrm{GW}\left(\psi'(\{0,1\}^{2rn})\right).$$

Since the dependence of our claimed bound on the Gaussian width is polynomial in n, the extra vertices will result in an extra factor depending only on d. It thus suffices to prove the theorem for the case where $H_1, \ldots, H_k$ are $2r$-uniform.

Observe that since the polynomials ψ_i are multilinear, the Gaussian width is bounded from above by replacing binary vectors with sign vectors. In particular,

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

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 p_{H_i} as in Lemma 2.2. Then, for every $x \in \{-1,1\}^n$,

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

where in the inequality we used that $x^{\otimes m}$ has Euclidean norm $\sqrt{N}$. Taking expectations, it then follows from Theorem 2.1 that the Gaussian width of $\psi(\{0,1\}^n)$ is at most

$$\frac{n}{c_r} \mathbb{E}\Big[\Big\| \sum_{i=1}^k g_i A_i \Big\|\Big] \lesssim \frac{n}{c_r} \sqrt{\log N} \Big(\sum_{i=1}^k \|A_i\|^2 \Big)^{1/2} \lesssim_r nt \sqrt{kn^{1-1/r} \log n},$$

where in the second inequality we used that $\|A_i\| \leq O_r(t)$ for each $i \in [k]$. □

3. Proof of the matrix lemma

In this section we prove Lemma 2.2. The starting point is a decomposition of a bounded-degree hypergraph into a small number of matchings. For this, we use the following basic result on edge colorings. 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.

Lemma 3.1. *Let H be a d-hypergraph. Then,*

$$\Delta(H) \leq \chi_E(H) \leq d(\Delta(H) - 1) + 1.$$

Proof: Clearly $\chi_E(H) \geq \Delta(H)$ 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 $d(\Delta(H)-1)$ other edges, $\Delta(G) \le d(\Delta(H)-1)$. □

To deal with matchings, we introduce the following definitions. Let $\mathcal{M} \subseteq \binom{[n]}{2r}$ be a maximal $2r$-matching of $[n]$. Let $s = 200 \cdot 4^r$. 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 $i \in [m] \smallsetminus I$, 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

$$(x^{\otimes m})_f (x^{\otimes m})_g = \prod_{i=1}^{m} x_{f(i)} x_{g(i)} = \prod_{j \in S} x_j. \tag{4}$$

Define the set of ordered pairs

$$\mathcal{P} = \big\{(f,g) : f \text{ is } s\text{-good and } g \text{ complements } f\big\}. \tag{5}$$

Proposition 3.2. *Let $\mathcal{P}$ be as in (5). 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. □

Proposition 3.3. *For every $(f,g) \in \mathcal{P}$, we have that g is s^2-good.*

Proof: Let $S \in \mathcal{M}$ and $(f,g) \in \mathcal{P}$ be such that (f,g) covers S. Consider the histograms $F, G : [n] \to \{0,1,\dots,m\}$ 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 s-good for (f, g) to belong to $\mathcal{P}$. It follows that

$$\phi(g) = \sum_{S' \in \mathcal{M}} \mu_{S'}(g) \leq 4^r + 2^r \sum_{S' \in \mathcal{M}} \mu_{S'}(f) = 4^r + 2^r \phi(f) \leq s^2,$$

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

Lemma 3.4 (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 $m = C_r n^{1-1/r}$,*

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

We postpone the proof of Lemma 3.4 to Section 4.

Corollary 3.5. *Let $\mathcal{P}$ be as in* (5) *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 $s^2(r!)$ ones.*

Proof: The first claim follows from Lemma 3.4 and the fact that $|\mathcal{P}|$ is at least the number of s-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 $s(r!)$ ones and by Proposition 3.3, every column of A has at most $s^2(r!)$ ones. □

With this, we can now prove Lemma 2.2.

Proof of Lemma 2.2: Let $t = \Delta(H)$. By Lemma 3.1, H can be decomposed into $\chi_E(H) \leq 2rt$ 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 (5) 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 (4) and Proposition 3.2 that for each $x \in \{-1, 1\}^n$, we have

$$\Big\langle \sum_{i=1}^{\chi_E(H)} (A_i + A_i^{\mathsf{T}}) x^{\otimes m}, x^{\otimes m} \Big\rangle = 2 \sum_{i=1}^{\chi_E(H)} \frac{|\mathcal{P}_i|}{|\mathcal{M}_i|} \sum_{S \in \mathcal{F}_i} \prod_{j \in S} x_i. \tag{6}$$

Since all $\mathcal{M}_i$ are maximal, they have the same size, as do the $\mathcal{P}_i$. Hence, by Corollary 3.5, there exists a constant $c_r \in (0, 1]$ such that the right-hand side of (6) equals $(2c_r N/n) p_H(x)$.

Let G be the graph with adjacency matrix A, allowing for parallel edges. Then G has degree at most $2ts^2(r!)$. It follows from Lemma 3.1 that G can be partitioned into $O_r(t)$ matchings. Since the adjacency matrix of a matching has unit norm, we get that $\|A\| \leq O_r(t)$. □

4. Proof of the generalized birthday paradox.

For the proof of Lemma 3.4, 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

$$\Pr[Y = \ell] = \frac{e^{-\mu}\mu^\ell}{\ell!}, \qquad \forall \ell = 0, 1, 2, \ldots \tag{7}$$

Proposition 4.1. *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 4.2. *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 3.4: 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 4.2, 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 4.2. 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 4.2, we have

$$\begin{aligned} \Pr[\phi(h) = 0] &= \mathbb{E}[\psi(\mathbf{X})] \\ &\leq 2\mathbb{E}[\psi(\mathbf{Y})] \\ &= 2 \prod_{S\in\mathcal{M}} \mathbb{E}\Big[\prod_{T\in\binom{S}{r}} \Big(1 - \prod_{i\in T} 1_{\geq 1}(Y_i)\Big)\Big], \end{aligned} \tag{8}$$

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 (8) 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 (8) give

$$\Pr[\phi(h) = 0] \leq 2\exp\left(-\frac{C_r^r|\mathcal{M}|}{en}\right) \leq 2\exp\left(-\frac{C_r^r\lfloor n/r \rfloor}{en}\right) \leq 2\exp\left(-\frac{C_r^r}{2er}\right). \tag{9}$$

Set $C_r = (6er)^{1/r}$, then the above right-hand side is at most 1/4. Next, we upper bound the probability that $\phi(h) \geq s = 200 \cdot 4^r$. 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 4.2 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 (9), we get that h is s-good with probability at least 1/2. □

5. Random differences in Szemerédi's Theorem

In this section we prove Theorem 1.3. We 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 $\mathbb{Z}/N\mathbb{Z}$. 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 $1 - o(1)$, every subset $A \subseteq \mathbb{Z}/N\mathbb{Z}$ of size at least αN contains a proper arithmetic progression of length $\ell+1$ with common difference in D.*

Proof: We will arrive at a contradiction assuming that the statement is false. Let $\Gamma = \mathbb{Z}/N\mathbb{Z}$. For $f : \Gamma \to \mathbb{R}$ and $y \in \Gamma \setminus \{0\}$, 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 $(f(x))_{x \in \Gamma}$. For a multiset $S \subseteq \Gamma \setminus \{0\}$, define

$$\Lambda_S(f) = \frac{1}{|S|}\sum_{y \in S} \phi_y(f).$$

If $f = 1_A$, then this counts the fraction of proper $(\ell+1)$-term APs with common difference in S that lie completely in A. Note that $\mathbb{E}_D[\Lambda_D(f)] = \Lambda_{\Gamma\setminus\{0\}}(f)$.

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 $A \subseteq \Gamma$ of size at least αN with no proper $(\ell+1)$-term APs whose common difference lies in D. Then,

$$\Pr_D\Big[\inf_{A:|A|\geq \alpha N} \Lambda_D(1_A) = 0\Big] = \Omega(1).$$

By Proposition 5.1, for every $A \subseteq \Gamma$ of size at least αN, we have that $\Lambda_{\Gamma\setminus\{0\}}(1_A) \geq \epsilon$. We are going to apply a standard symmetrization trick to establish a connection with Gaussian width. Let D' be an independent copy of D. Then,

$$\begin{aligned}
\epsilon &\lesssim \mathbb{E}_D\Big[\sup_{A:|A|\geq\alpha N} \big|\Lambda_D(1_A) - \Lambda_{\Gamma\setminus\{0\}}(1_A)\big|\Big] \\
&= \mathbb{E}_D\Big[\sup_{A:|A|\geq\alpha N} \big|\Lambda_D(1_A) - \mathbb{E}_{D'}[\Lambda_{D'}(1_A)]\big|\Big] \\
&\leq \mathbb{E}_{D,D'}\Big[\sup_{A:|A|\geq\alpha N} \big|\Lambda_D(1_A) - \Lambda_{D'}(1_A)\big|\Big] \\
&= \mathbb{E}_{y_1,\ldots,y_k,y_1',\ldots,y_k'\in\Gamma\setminus\{0\}}\Big[\sup_{A:|A|\geq\alpha N} \Big|\frac{1}{k}\sum_{i=1}^{k} \phi_{y_i}(1_A) - \phi_{y_i'}(1_A)\Big|\Big]
\end{aligned}$$

Observe that for i.i.d. random $y, y' \in \Gamma\setminus\{0\}$, the random variable $\phi_y(1_A) - \phi_{y'}(1_A)$ is symmetric in the sense that it has the same distribution as its negation. Let $\sigma_1,\ldots,\sigma_k$ be independent uniformly distributed $\{-1,1\}$-valued random variables. Then it follows from the above that

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

Let us fix $y_1,\ldots,y_k \in \Gamma\setminus\{0\}$. Each ϕ_{y_i} can be written as $\phi_{y_i} = N^{-1}p_{H_i}$ (as in (3)) 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. Let $g_1,\ldots,g_k$ be independent $N(0,1)$ random variables. Then we can bound

$$\begin{aligned}\mathbb{E}_\sigma\Big[\sup_{A:|A|\geq\alpha N}\Big|\frac{1}{k}\sum_{i=1}^k \sigma_i\phi_{y_i}(1_A)\Big|\Big] &\lesssim \frac{1}{k}\mathbb{E}_g\Big[\sup_A\Big|\sum_{i=1}^k g_i\phi_{y_i}(1_A)\Big|\Big]\\ &= \frac{1}{Nk}\mathbb{E}_g\Big[\sup_A\Big|\sum_{i=1}^k g_i p_{H_i}(1_A)\Big|\Big]\\ &\lesssim_\ell \frac{1}{k}\sqrt{kN^{1-1/\lceil(\ell+1)/2\rceil}\log N},\end{aligned}$$

where the last line follows directly from Theorem 1.1. Thus we get $k \lesssim_\ell N^{1-1/\lceil(\ell+1)/2\rceil}\log N$ which is a contradiction. □

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:

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

□

Proof of Theorem 1.3: Let D_k be a random subset of $\mathbb{Z}/N\mathbb{Z}\smallsetminus\{0\}$ 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 $1-o(1)$, then D_p will also be ℓ-intersective with probability $1-o(1)$ 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,

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

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 $o(1)$. 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. Upper tails for arithmetic progressions in random sets

Here we prove Theorem 1.4. Let $\Gamma=\mathbb{Z}/N\mathbb{Z}$. In the following we identify maps from a set S to $\mathbb{R}$ with vectors in $\mathbb{R}^S$. For $f:\Gamma\to\mathbb{R}$, define

$$\Lambda_k(f) = \sum_{a,b\in\Gamma,b\neq 0} f(a)f(a+b)f(a+2b)\cdots f(a+(k-1)b). \tag{10}$$

Observe that for a subset $A \subseteq \Gamma$, we have that $\Lambda_k(1_A)$ counts the number of proper k-term arithmetic progressions in A. Moreover, Λ_k is an N-variate polynomial of degree k. Recall that the gradient of a polynomial $p \in \mathbb{R}[x_1, \dots, x_n]$ is the mapping $\nabla p : \mathbb{R}^n \to \mathbb{R}^n$ whose ith coordinate is given by $(\nabla p)_i = (\partial p/\partial x_i)(x)$. The proof of Theorem 1.4 follows from a simple corollary of Theorem 1.1 and one of the main results of [BGSZ18]. For the corollary, we consider polynomial mappings given by gradients of polynomials of the form (3).

Corollary 6.1. *Let n, t, d be positive integers. Let $H = ([n], E)$ be a $(d+1)$-hypergraph such that at most t edges are incident on any given pair of vertices. Then,*

$$\frac{1}{n}\,\mathrm{GW}\left((\nabla p_H)(\{0,1\}^n)\right) \lesssim_d t n^{1-\frac{1}{2\lceil d/2\rceil}}\sqrt{\log n}.$$

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

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

The claim now follows from Theorem 1.1 as $p_{H_i} = (\nabla p_H)_i$ each H_i has degree at most t. □

Theorem 6.2 (Bhattacharya–Ganguly–Shao–Zhao)**.** *Let $k \geq 3$ be a fixed integer and let σ, τ be positive real numbers such that*

$$\frac{1}{N}\,\mathrm{GW}\left(\nabla\Lambda_k(\{0,1\}^\Gamma)\right) \lesssim N^{1-\sigma}(\log N)^\tau.$$

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,

$$\log \Pr[\Lambda_k(\Gamma_p) \geq (1+\delta)\mathbb{E}\Lambda_k(\Gamma_p)] = -\big(1+o(1)\big)\,\phi_p\big((1+o(1))\delta\big). \tag{11}$$

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

$$\phi_p(\delta) \asymp N\,\min\{\sqrt{\delta}p^{k/2}\log(1/p), \delta^2 p\}.$$

Proof of Theorem 1.4: Let $H = (\Gamma, E)$ be the hypergraph whose edges are the (unordered) proper k-term arithmetic progressions in Γ. Then, accounting for the fact that Λ_k distinguishes between the same progression with step b run forward from a point a or backward from $a+(k-1)b$ and since N is prime, we have $2p_H = \Lambda_k$. We claim that every pair of distinct vertices appears in $O(k^2)$ edges. First note that H is 2-transitive, since for any two pairs of distinct vertices $(a,b),(c,d)$, the affine linear map $x \mapsto c(x-b)/(a-b) + d(x-a)/(b-a)$ sends a to c, b to d and preserves progressions. It follows that every pair of distinct vertices is contained in the same number of edges. Since each edge contains $\binom{k}{2}$ pairs, the claim follows by double-counting. By Corollary 6.1, we may thus set $\sigma = 1/(2\lceil (k-1)/2\rceil)$ and $\tau = 1/2$ in Theorem 6.2 and it follows that for constant δ, the estimate (11) holds if

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

Taking kth roots now gives the claim. □

7. Proof of Lemma 1.5

In this section we give a proof Lemma 1.5. As explained in the proof of Theorem 1.1, it suffices to prove the statement when the coordinates of ψ are given by p_{H_i} (as in (3)) for d-uniform hypergraphs $H_1, \ldots, H_k$. Let Λ_{H_i} be a d-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,

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

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,

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

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 $\|\Lambda_{H_i}\| \leq 1$. If not, by Lemma 3.1, we can decompose H_i into $d\Delta(H_i)$ matchings and use triangle inequality to conclude that $\|\Lambda_{H_i}\| \leq d\Delta(H_i)$ which gives the desired bound.

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.

References

[BDG17] Jop Briët, Zeev Dvir, and Sivakanth Gopi. Outlaw Distributions and Locally Decodable Codes. In *8th Innovations in Theoretical Computer Science Conference (ITCS 2017)*, volume 67 of *Leibniz*

International Proceedings in Informatics (LIPIcs), pages 20:1–20:19. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2017.

[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.

[BGSZ18] Bhaswar Bhattacharya, Shirshendu Ganguly, Xuancheng Shao, and Yufei Zhao. Upper tails for arithmetic progressions in a random set. *International Mathematics Research Notices*, 2018. To appear. Available at arXiv preprint: 1605.02994.

[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.

[Tal14] Michel Talagrand. *Upper and lower bounds for stochastic processes*, volume 60 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]*. Springer, Heidelberg, 2014. Modern methods and classical problems.

[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