On the Threshold for Szemerédi's Theorem with Random Differences

Jop Briët Davi Castro-Silva

Submitted: Sep 28, 2023; Accepted: Jul 9, 2024; Published: Oct 4, 2024

Abstract

Using recent developments on the theory of locally decodable codes, we prove that the critical size for Szemerédi's theorem with random differences is bounded from above by $N^{1-\frac{2}{k}+o(1)}$ for length-k progressions. This improves the previous best bounds of $N^{1-\frac{1}{\lceil k/2 \rceil}+o(1)}$ for all odd k.

Mathematics Subject Classifications: 11B30, 05D40

1 Introduction

Szemerédi [20] proved that dense sets of integers contain arbitrarily long arithmetic progressions, a result which has become a hallmark of additive combinatorics. Multiple proofs of this result were found over the years, using ideas from combinatorics, ergodic theory and Fourier analysis over finite abelian groups.

Furstenberg's ergodic theoretic proof [14] opened the floodgates to a series of powerful generalizations. In particular, it led to versions of Szemerédi's theorem where the common differences for the arithmetic progressions are restricted to very sparse sets. We say that a set $D \subseteq [N]$ is *t-intersective* if any positive-density set $A \subseteq [N]$ contains an $(t+1)$-term arithmetic progression with common difference in D. Szemerédi's theorem implies that for large enough N_0, the set $\{0, 1, \ldots, N_0\}$ is t-intersective for $N \geqslant N_0$. Non-trivial examples include a result of Bergelson and Leibman [3] showing that the perfect squares (and more generally, images of integer polynomials with zero constant term) are t-intersective for every t, and a special case of a result of Wooley and Ziegler [23] showing the same for the prime numbers minus one.

The existence of such sparse intersective sets motivated the problem of showing whether, in fact, random sparse sets are typically intersective. The task of making this quantitative falls within the scope of research on threshold phenomena. We say that a property of subsets of $[N]$, given by a family $\mathcal{F} \subseteq 2^{[N]}$, is *monotone* if $A \in \mathcal{F}$ and

CWI & QuSoft, Science Park 123, 1098 XG Amsterdam, The Netherlands (j.briet@cwi.nl, davisilva15@gmail.com).

$A \subseteq B \subseteq [N]$ imply $B \in \mathcal{F}$. The *critical size* $m^* = m^*(N)$ of a property is the least m such that a uniformly random m-element subset of $[N]$ has the property with probability at least $1/2$. (This value exists if $\mathcal{F}$ is non-empty and monotone, as this probability then increases monotonically with m). A famous result of Bollobás and Thomason [4] asserts that every monotone property has a threshold function; this is to say that the probability

$$p(m) = \Pr_{A \in \binom{[N]}{m}}[A \in \mathcal{F}]$$

spikes from $o(1)$ to $1 - o(1)$ when m increases from $o(m^*)$ to $\omega(m^*)$. In general, it is notoriously hard to determine the critical size of a monotone property.

This problem is also wide open for the property of being t-intersective, which is clearly monotone, and for which we denote the critical size by $m_t^*(N)$. Bourgain [5] showed that the critical size for 1-intersective sets is given by $m_1^*(N) \asymp \log N$; at present, this is the only case where precise bounds are known. It has been conjectured [13] that $\log N$ is the correct bound for all fixed t, and indeed no better lower bounds are known for $t \geqslant 2$. It was shown by Frantzikinakis, Lesigne and Wierdl [12] and independently by Christ [11] that

$$m_2^*(N) \ll N^{\frac{1}{2}+o(1)}. \tag{1}$$

The same upper bound was later shown to hold for $m_3^*(N)$ by the first author, Dvir and Gopi [7]. More generally, they showed that

$$m_t^*(N) \ll N^{1-\frac{1}{\lceil (t+1)/2 \rceil}+o(1)}, \tag{2}$$

which improved on prior known bounds for all $t \geqslant 3$. The appearance of the ceiling function in these bounds is due to a reduction for even t to the case $t+1$. The reason for this reduction originates from work on locally decodable error correcting codes [16]. It was shown in [7] that lower bounds on the block length of $(t+1)$-query locally decodable codes (LDCs) imply upper bounds on m_t^*. The bounds (2) then followed directly from the best known LDC bounds; see [8] for a direct proof of (2), however.

For the same reason, a recent breakthrough of Alrabiah et al. [1] on 3-query LDCs immediately implies an improvement of (1) to

$$m_2^*(N) \ll N^{\frac{1}{3}+o(1)}.$$

For technical reasons, their techniques do not directly generalize to improve the bounds for q-query LDCs with $q \geqslant 4$. Here, we use the ideas of [1] to directly prove upper bounds on m_t^*. Due to the additional arithmetic structure in our problem, it is possible to simplify the exposition and, more importantly, apply the techniques to improve the previous best known bounds for all even $t \geqslant 2$.

Theorem 1. *For every integer $t \geqslant 2$, we have that*

$$m_t^*(N) \ll N^{1-\frac{2}{t+1}+o(1)}.$$

The arguments presented here in fact work in greater generality, and hold for any finite additive group G whose size is coprime to $t!$ (so as not to incur in divisibility issues when considering $(t+1)$-term arithmetic progressions).

Let G be a finite additive group, $t \geqslant 1$ be an integer and $\varepsilon \in (0,1)$. We say that a set $S \subseteq G$ is (t,ε)-*intersective* if every subset $A \subseteq G$ of size $|A| \geqslant \varepsilon|G|$ contains an $(t+1)$-term arithmetic progression with common difference in D. We denote the critical size for the property of being (t,ε)-intersective in G by $m^*_{t,\varepsilon}(G)$. Our main result is the following:

Theorem 2. *For every $t \geqslant 2$ and $\varepsilon \in (0,1)$, there exists $C(t,\varepsilon) > 0$ such that*

$$m^*_{t,\varepsilon}(G) \leqslant C(t,\varepsilon)(\log|G|)^{2t+3}|G|^{1-\frac{2}{t+1}}$$

for every additive group G whose size is coprime to $t!$.

Note that Theorem 1 follows easily from this last result by embedding $[N]$ into a group of the form $\mathbb{Z}/p\mathbb{Z}$, where p is a prime between $(t+1)N$ and $2(t+1)N$. We omit the standard details.

2 Preliminaries

2.1 Notation

We write $\biguplus$ for a disjoint union. Our (standard) asymptotic notation is defined as follows. Given a parameter n which grows without bounds and a function $f : \mathbb{R}_+ \to \mathbb{R}_+$, we write: $g(n) = o(f(n))$ to mean $g(n)/f(n) \to 0$; $g(n) = \omega(f(n))$ to mean $g(n)/f(n) \to \infty$; $g(n) \ll f(n)$ to mean that $g(n) \leqslant Cf(n)$ holds for some constant $C > 0$ and all n; and $g(n) \asymp f(n)$ to mean both $g(n) \ll f(n)$ and $f(n) \ll g(n)$. When the implied constant in the asymptotics depends on some parameter (say ε), we indicate this by adding said parameter as a subscript in the asymptotic notation (replacing $\ll$ by $\ll_\varepsilon$, say).

2.2 Matrix norms and inequalities

Our arguments will rely heavily on the analysis of high-dimensional matrices. Here we recall the matrix inequalities which will be needed.

If $M \in \mathbb{R}^{d\times d}$ is a matrix, we define its operator norms

$$\begin{aligned}\|M\|_2 &= \max\left\{u^TMv :\ \|u\|_2 = \|v\|_2 = 1\right\}\\ \|M\|_{\infty\to 1} &= \max\left\{u^TMv :\ \|u\|_\infty = \|v\|_\infty = 1\right\}\\ \|M\|_{1\to 1} &= \max\left\{u^TMv :\ \|u\|_\infty = \|v\|_1 = 1\right\}.\end{aligned}$$

We will make use of the following simple inequalities:

$$\|M\|_{\infty\to 1} \leqslant d\|M\|_2, \quad \|M\|_{\infty\to 1} \leqslant \sum_{i=1}^{d}\|M(i,\cdot)\|_1$$

and, when M is symmetric,

$$\|M\|_2 \leqslant \|M\|_{1\to 1}.$$

We will also use the following noncommutative version of Khintchine's inequality, which can be extracted from a result of Tomczak-Jaegermann [21]:

Theorem 3. *Let $n, d \geqslant 1$ be integers, and let $A_1, \dots, A_n$ be any sequence of $d \times d$ real matrices. Then*

$$\mathbb{E}_{\sigma\in\{-1,1\}^n} \left\| \sum_{i=1}^n \sigma_i A_i \right\|_2 \leqslant 10\sqrt{\log d} \left(\sum_{i=1}^n \|A_i\|_2^2 \right)^{1/2}.$$

2.3 Polynomial concentration bounds

We will need a well-known concentration inequality for polynomials due to Kim and Vu [17], which requires the introduction of some extra notation. Let $H = (V, E)$ be a hypergraph, where we allow for repeated edges (so E may be a multiset), and let $f : \{0,1\}^V \to \mathbb{R}$ be the polynomial given by

$$f(x) = \sum_{e\in E} \prod_{v\in e} x_v. \tag{3}$$

For a set $A \subseteq V$, define

$$f_A(x) = \sum_{e\in E:\, A\subseteq e} \prod_{v\in e\setminus A} x_v,$$

where the monomial corresponding to the empty set is defined to be 1. For $p \in (0,1)$, we say that X is a *p-Bernoulli random variable on $\{0,1\}^V$*, denoted $X \sim \mathrm{Bern}(p)^V$, if its coordinates are all independent and each equals 1 with probability p (and equals 0 with probability $1-p$). For each $i \in \{0, 1, \dots, |V|\}$, define

$$\mu_i = \max_{A\in\binom{V}{i}} \mathbb{E}_{X\sim\mathrm{Bern}(p)^V} f_A(X).$$

Note that μ_0 is just the expectation of $f(X)$. Define also the quantities

$$\mu = \max_{i\in\{0,1,\dots,|V|\}} \mu_i \qquad \text{and} \qquad \mu' = \max_{i\in\{1,2,\dots,|V|\}} \mu_i.$$

The polynomial concentration inequality of Kim and Vu is given as follows:

Theorem 4. *For every $k \in \mathbb{N}$, there exist constants $C, C' > 0$ such that the following holds. Let $H = (V, E)$ be an n-vertex hypergraph whose edges have size at most k, and let f be given by (3). Then, for any $\lambda > 1$, we have*

$$\Pr\big[|f(X) - \mu_0| > C\lambda^{k-\frac{1}{2}}\sqrt{\mu\mu'}\big] \leqslant C' \exp\big(-\lambda + (k-1)\log n\big).$$

To suit our needs, we will use a slight variant of this result, which follows easily from it and the following basic proposition.

Proposition 5. *Let $f : \{0,1\}^n \to \mathbb{R}_+$ be a monotone increasing function and $p \in (\frac{16}{n}, 1)$. Then, for any integer $0 \leqslant t \leqslant pn/2$,*

$$\mathbb{E}_{S \in \binom{[n]}{t}} f(1_S) \leqslant \frac{1}{2}\, \mathbb{E}_{X \sim \mathrm{Bern}(p)^n} f(X).$$

Proof. By direct calculation,

$$\begin{aligned}
\mathbb{E}_{X \sim \mathrm{Bern}(p)^n} f(X) &= \sum_{i=0}^{n} p^i (1-p)^{n-i} \sum_{S \in \binom{[n]}{i}} f(1_S) \\
&= \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i}\, \mathbb{E}_{S \in \binom{[n]}{i}} f(1_S) \\
&\geqslant \sum_{i \geqslant t} \binom{n}{i} p^i (1-p)^{n-i}\, \mathbb{E}_{S \in \binom{[n]}{t}} f(1_S) \\
&\geqslant \frac{1}{2}\, \mathbb{E}_{S \in \binom{[n]}{t}} f(1_S),
\end{aligned}$$

where in the third line we used monotonicity of f and the fourth line follows from the Chernoff bound. □

Corollary 6. *For every $k \in \mathbb{N}$, there exist constants $C, C' > 0$ such that the following holds. Let $H = (V, E)$ be an n-vertex hypergraph whose edges have size at most k, let f be given as in (3) and let $p \in (\frac{16}{n}, 1)$. Then, for any integer $0 \leqslant t \leqslant pn/2$, we have*

$$\Pr_{S \in \binom{V}{t}} \big[f(1_S) \geqslant C (\log n)^{k - \frac{1}{2}} \mu \big] \leqslant \frac{C'}{n^4}.$$

Proof. For a sufficiently large constant $C = C(k) > 0$, let $g : \{0,1\}^n \to \{0,1\}$ be the indicator function

$$g(1_S) = \mathbf{1}\big[f(1_S) \geqslant C (\log n)^{k - \frac{1}{2}} \mu \big].$$

Since f is monotone, so is g. Setting $\lambda = (3+k) \log n$, it follows from Theorem 4 that

$$\mathbb{E}_{X \sim \mathrm{Bern}(p)^n}\, g(X) \leqslant \frac{C'}{n^4}.$$

The result now follows from Proposition 5. □

3 The main argument

In this section we will prove Theorem 2, our main result. It will be more convenient to shift attention from the degree t of intersectivity to the length $k := t+1$ of the associated arithmetic progressions.

Fix then an integer $k \geqslant 3$ for the length of the progressions and a positive parameter $\varepsilon > 0$. Let G be an additive group with N elements, where N is coprime to $(k-1)!$ and

is assumed to be sufficiently large relative to k and ε for our arguments to hold. Recall that we wish to show that $m^*_{k-1,\varepsilon}(G) \ll_{k,\varepsilon} (\log N)^{2k+1} N^{1-\frac{2}{k}}$.

Instead of considering random intersective sets, it will be simpler to consider random *intersective sequences*, where a sequence in G^m is $(k-1,\varepsilon)$-intersective if the set of its distinct elements is. Clearly, the probability that a uniformly random m-element sequence is $(k-1,\varepsilon)$-intersective is at most the probability that a uniform m-element set is. Since we are interested in proving upper bounds on the critical size, it suffices to bound the minimal m such that a random sequence in G^m is $(k-1,\varepsilon)$-intersective with probability at least $1/2$.

3.1 Reducing to an inequality about sign averages

Given a sequence of differences $D = (d_1, \dots, d_m) \in G^m$ and some set $A \subseteq G$, let $\Lambda_D(A)$ be the normalized count of k-APs with common difference in D which are contained in A:

$$\Lambda_D(A) = \mathbb{E}_{i\in[m]}\mathbb{E}_{x\in G}\prod_{\ell=0}^{k-1} A(x+\ell d_i).$$

Similarly, we denote by $\Lambda_G(A)$ the proportion of all k-APs which are contained in A:

$$\Lambda_G(A) = \mathbb{E}_{d\in G}\mathbb{E}_{x\in G}\prod_{\ell=0}^{k-1} A(x+\ell d).$$

By a suitable generalization of Szemerédi's theorem, we know that

$$\Lambda_G(A) \gg_{k,\varepsilon} 1 \quad \text{for all } A \subseteq G \text{ with } |A| \geqslant \varepsilon N. \tag{4}$$

This can be proven, for instance, by using the *hypergraph removal lemma* of Gowers [15] and Nagle, Rödl, Schacht and Skokan [19, 18].

Now suppose $m \in [N]$ is an integer for which

$$\Pr_{D\in G^m}\big(\exists A \subseteq G :\ |A| \geqslant \varepsilon N,\ \Lambda_D(A) = 0\big) \geqslant 1/2. \tag{5}$$

Noting that $\mathbb{E}_{D'\in G^m}\Lambda_{D'}(A) = \Lambda_G(A)$, by combining inequalities (5) and (4) we conclude that

$$\mathbb{E}_{D\in G^m} \max_{A\subseteq G:\,|A|\geqslant\varepsilon N} \big|\Lambda_D(A) - \mathbb{E}_{D'\in G^m}\Lambda_{D'}(A)\big| \gg_{k,\varepsilon} 1.$$

Below, we will no longer need the condition that $|A| \geqslant \varepsilon N$ in maxima over $A \subseteq G$. This positive density assumption is only used through (4).

We next apply a simple symmetrization argument given in [8, page 8690] to write this in a more convenient form:

Lemma 7 (Symmetrization)**.** *Let $c > 0$, and suppose that*

$$\mathbb{E}_{D\in G^m} \max_{A\subseteq G} \big|\Lambda_D(A) - \mathbb{E}_{D'\in G^m}\Lambda_{D'}(A)\big| \geqslant c.$$

Then

$$\mathbb{E}_{D\in G^m}\mathbb{E}_{\sigma\in\{-1,1\}^m} \max_{A\subseteq G} \left|\mathbb{E}_{i\in[m]}\mathbb{E}_{x\in G}\,\sigma_i \prod_{\ell=0}^{k-1} A(x+\ell d_i)\right| \geqslant \frac{c}{2}.$$

Proof. With slight abuse of notation, for each $d \in G$ and $A \subseteq G$, denote

$$\Lambda_d(A) = \mathbb{E}_{x\in G} \prod_{\ell=0}^{k-1} A(x+\ell d).$$

Then, by the triangle inequality, the first expression in Lemma 7 is bounded from above by

$$\mathbb{E}_{D,D'\in G^m} \max_{A\subseteq G} \Big|\frac{1}{m}\sum_{i=1}^{m} \big(\Lambda_{d_i}(A) - \Lambda_{d_i'}(A)\big)\Big|,$$

where D and D' are independent and uniformly distributed. Since the terms $\Lambda_{d_i}(A) - \Lambda_{d_i'}(A)$ are independent and symmetrically distributed, this expectation is unchanged if each of these terms is multiplied by an arbitrary sign. In particular, this expectation equals

$$\mathbb{E}_{D,D'\in G^m}\mathbb{E}_{\sigma\in\{-1,1\}^m} \max_{A\subseteq G} \Big|\frac{1}{m}\sum_{i=1}^{m} \sigma_i\big(\Lambda_{d_i}(A) - \Lambda_{d_i'}(A)\big)\Big|.$$

Using the triangle inequality again, and the fact that D and D' have the same distribution, we get that this is bounded from above by

$$2\mathbb{E}_{D\in G^m}\mathbb{E}_{\sigma\in\{-1,1\}^m} \max_{A\subseteq G} \Big|\frac{1}{m}\sum_{i=1}^{m} \sigma_i\Lambda_{d_i}(A)\Big|.$$

This proves the result. □

The appearance of the expectation over signs $\sigma \in \{-1,1\}^m$ is crucial to our arguments. By an easy multilinearity argument, we can replace the set $A \subseteq G$ (which can be seen as a vector in $\{0,1\}^G$) by a vector $Z \in \{-1,1\}^G$. In combination with (5) and Lemma 7, this gives

$$\mathbb{E}_{D\in G^m}\mathbb{E}_{\sigma\in\{-1,1\}^m} \max_{Z\in\{-1,1\}^G} \Big|\mathbb{E}_{i\in[m]}\mathbb{E}_{x\in G}\, \sigma_i \prod_{\ell=0}^{k-1} Z(x+\ell d_i)\Big| \gg_{k,\varepsilon} 1. \tag{6}$$

The change from $\{0,1\}^G$ to $\{-1,1\}^G$ is a convenient technicality so that we can ignore terms which get squared in a product.

3.2 Dealing with an odd number of terms

The last inequality (6) is what we need to prove the result for even values of k using the arguments we will outline below. For odd values of k, however, this inequality is unsuited due to the odd number of factors inside the product. The main idea from [1] to deal with this case is to apply a "Cauchy-Schwarz trick" to obtain a better suited inequality:

Lemma 8 (Cauchy-Schwarz trick)**.** *Let $c > 0$, and suppose $m \geqslant 2/c^2$ is an integer for which*

$$\mathbb{E}_{D\in G^m}\mathbb{E}_{\sigma\in\{-1,1\}^m}\max_{Z\in\{-1,1\}^G}\left|\mathbb{E}_{i\in[m]}\mathbb{E}_{x\in G}\,\sigma_i\prod_{\ell=0}^{k-1}Z(x+\ell d_i)\right|\geqslant c.$$

Then there exists a partition $[m] = L \uplus R$ such that

$$\mathbb{E}_{D\in G^m}\mathbb{E}_{\substack{\sigma\in\{-1,1\}^L\\ \tau\in\{-1,1\}^R}}\max_{Z\in\{-1,1\}^G}\sum_{\substack{i\in L\\ j\in R}}\sum_{x\in G}\sigma_i\tau_j\prod_{\ell=1}^{k-1}Z(x+\ell d_i)Z(x+\ell d_j)\geqslant\frac{c^2m^2N}{8}.$$

Proof. By Cauchy-Schwarz, for any $Z\in\{-1,1\}^G$ we have

$$\begin{aligned}\left|\mathbb{E}_{i\in[m]}\mathbb{E}_{x\in G}\,\sigma_i\prod_{\ell=0}^{k-1}Z(x+\ell d_i)\right|^2&=\left|\mathbb{E}_{x\in G}\,Z(x)\cdot\left(\mathbb{E}_{i\in[m]}\sigma_i\prod_{\ell=1}^{k-1}Z(x+\ell d_i)\right)\right|^2\\&\leqslant\big(\mathbb{E}_{x\in G}\,Z(x)^2\big)\mathbb{E}_{x\in G}\left(\mathbb{E}_{i\in[m]}\sigma_i\prod_{\ell=1}^{k-1}Z(x+\ell d_i)\right)^2\\&=\mathbb{E}_{x\in G}\mathbb{E}_{i,j\in[m]}\,\sigma_i\sigma_j\prod_{\ell=1}^{k-1}Z(x+\ell d_i)Z(x+\ell d_j).\end{aligned}$$

Applying Cauchy-Schwarz again, we conclude from our assumption that

$$\begin{aligned}c^2&\leqslant\mathbb{E}_{D\in G^m}\mathbb{E}_{\sigma\in\{-1,1\}^m}\max_{Z\in\{-1,1\}^G}\left|\mathbb{E}_{i\in[m]}\mathbb{E}_{x\in G}\,\sigma_i\prod_{\ell=0}^{k-1}Z(x+\ell d_i)\right|^2\\&\leqslant\mathbb{E}_{D\in G^m}\mathbb{E}_{\sigma\in\{-1,1\}^m}\max_{Z\in\{-1,1\}^G}\mathbb{E}_{x\in G}\mathbb{E}_{i,j\in[m]}\,\sigma_i\sigma_j\prod_{\ell=1}^{k-1}Z(x+\ell d_i)Z(x+\ell d_j).\end{aligned}$$

Now consider a uniformly random partition $[m] = L \uplus R$, so that for any $i, j \in [m]$ with $i \neq j$ we have $\Pr_{L,R}(i \in L,\, j \in R) = 1/4$; then

$$\begin{aligned}&\mathbb{E}_{i,j\in[m]}\,\sigma_i\sigma_j\prod_{\ell=1}^{k-1}Z(x+\ell d_i)Z(x+\ell d_j)\\&\quad=\frac{1}{m^2}\sum_{\substack{i,j=1\\ i\neq j}}^{m}\sigma_i\sigma_j\prod_{\ell=1}^{k-1}Z(x+\ell d_i)Z(x+\ell d_j)+\frac{1}{m^2}\sum_{i=1}^{m}\sigma_i^2\prod_{\ell=1}^{k-1}Z(x+\ell d_i)^2\\&\quad=\frac{4}{m^2}\mathbb{E}_{L,R}\sum_{i\in L,j\in R}\sigma_i\sigma_j\prod_{\ell=1}^{k-1}Z(x+\ell d_i)Z(x+\ell d_j)+\frac{1}{m}.\end{aligned}$$

It follows that

$$c^2\leqslant\frac{1}{m}+\frac{4}{m^2}\mathbb{E}_{L,R}\mathbb{E}_{D\in G^m}\mathbb{E}_{\sigma\in\{-1,1\}^m}\max_{Z\in\{-1,1\}^G}\mathbb{E}_{x\in G}\sum_{i\in L,j\in R}\sigma_i\sigma_j\prod_{\ell=1}^{k-1}Z(x+\ell d_i)Z(x+\ell d_j).$$

Using that $m \geqslant 2/c^2$, we conclude there exists a choice of partition $[m] = L \uplus R$ satisfying the conclusion of the lemma. □

From now on we assume that k is odd, and write $k = 2r + 1$.[1] For $i, j \in [m]$, denote $P_i(x) = \{x + d_i, x + 2d_i, \ldots, x + 2rd_i\}$ and $P_{ij}(x) = P_i(x) \cup P_j(x)$, where we hide the dependence on the difference set D for ease of notation. From inequality (6) and Lemma 8 we conclude that

$$\mathbb{E}_{D\in G^m}\mathbb{E}_{\substack{\sigma\in\{-1,1\}^L\\ \tau\in\{-1,1\}^R}} \max_{Z\in\{-1,1\}^G} \sum_{\substack{i\in L\\ j\in R}} \sum_{x\in G} \sigma_i\tau_j \prod_{y\in P_{ij}(x)} Z(y) \gg_{k,\varepsilon} m^2N, \tag{7}$$

where (L, R) is a suitable partition of the index set $[m]$ and we assume (without loss of generality) that m is sufficiently large depending on ε and k.

From inequality (7) it follows that we can fix a "good" set $D \in G^m$ satisfying

$$\mathbb{E}_{\substack{\sigma\in\{-1,1\}^L\\ \tau\in\{-1,1\}^R}} \max_{Z\in\{-1,1\}^G} \sum_{\substack{i\in L\\ j\in R}} \sigma_i\tau_j \sum_{x\in G} \prod_{y\in P_{ij}(x)} Z(y) \gg_{k,\varepsilon} m^2N \tag{8}$$

and for which we have the technical conditions

$$\big|\big\{i \in L, j \in R : \ |P_{ij}(0)| \neq 4r\big\}\big| \ll_{k,\varepsilon} m^2/N \quad \text{and} \tag{9}$$

$$\max_{x\neq 0} \sum_{i=1}^{m} \sum_{\ell=-2r}^{2r} \mathbf{1}\{\ell d_i = x\} \ll \log N, \tag{10}$$

which are needed to bound the probability of certain bad events later on. Indeed:

- Denote by $X(D)$ the expression on the left-hand side of (8) for a given sequence $D \in G^m$. Then $X(D) \leqslant m^2N$ always holds, while by equation (7) we have $\mathbb{E}_{D\in G^m} X(D) \geqslant c_{k,\varepsilon} m^2N$ for some constant $c_{k,\varepsilon} > 0$. It follows that

$$\Pr\big[X(D) \geqslant c_{k,\varepsilon} m^2N/2\big] > c_{k,\varepsilon}/2.$$

- For $\ell, \ell' \in [2r]$ and independent uniform $d_i, d_j \in G$, we have that

$$\Pr[\ell d_i = \ell' d_j] = 1/N.$$

  The expectation of the left-hand side of (9) (taken with respect to $D \in G^m$) is then

$$\sum_{\substack{i\in L\\ j\in R}} \Pr\big[|P_{ij}(0)| \neq 4r\big] \leqslant \sum_{\substack{i\in L\\ j\in R}} \sum_{\ell,\ell'=1}^{2r} \Pr[\ell d_i = \ell' d_j] \leqslant \frac{r^2m^2}{N},$$

  and thus by Markov's inequality

$$\Pr\left[\big|\big\{i \in L, j \in R : \ |P_{ij}(0)| \neq 4r\big\}\big| \leqslant \frac{4}{c_{k,\varepsilon}} \frac{r^2m^2}{N}\right] \geqslant 1 - \frac{c_{k,\varepsilon}}{4}.$$

[1] The even case is similar but simpler. We focus on the odd case here because this is where we get new bounds.

- For a fixed $x \neq 0$, the inner sum in (10) is an indicator random variable that equals 1 with probability $4r/N$. Since these random variables are independent for different $i \in [m]$, the Chernoff bound implies that

$$\Pr\Bigg[\sum_{i=1}^{m}\sum_{\ell=-2r}^{2r} \mathbf{1}\{\ell d_i = x\} \geqslant (1+\delta)\frac{4rm}{N}\Bigg] \leqslant \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^{4rm/N}.$$

  Setting $1+\delta = N\log N/(rm)$, the union bound over all $x \neq 0$ gives that

$$\Pr\Bigg[\max_{x\neq 0}\sum_{i=1}^{m}\sum_{\ell=-2r}^{2r} \mathbf{1}\{\ell d_i = x\} \leqslant 4\log N\Bigg] \geqslant 1 - \frac{1}{N^3} > 1 - \frac{c_{k,\varepsilon}}{4}.$$

By the union bound, all three conditions above will hold simultaneously with positive probability, as wished.

3.3 Reducing to a matrix inequality problem

The next key idea is to construct matrices M_{ij} for which the quantity

$$\mathbb{E}_{\substack{\sigma\in\{-1,1\}^L\\ \tau\in\{-1,1\}^R}}\Bigg\|\sum_{i\in L, j\in R}\sigma_i\tau_j M_{ij}\Bigg\|_{\infty\to 1} \tag{11}$$

is related to the expression on the left-hand side of inequality (8). The reason for doing so is that this allows us to use strong *matrix concentration inequalities*, which can be used to obtain a good upper bound on the expectation (11); this in turn translates to an upper bound on m as a function of N, which is our goal. Such uses of matrix inequalities go back to work of Ben-Aroya, Regev and de Wolf [2], in turn inspired by work of Kerenidis and de Wolf [16] (see also [10]).

The matrices we will construct are indexed by sets of a given size s, where (with hindsight) we choose $s = \lfloor N^{1-2/k}\rfloor$. Recall that $k = 2r+1$. For $i \in L$, $j \in R$, define the matrix $M_{ij} \in \mathbb{R}^{\binom{G}{s}\times\binom{G}{s}}$ by

$$M_{ij}(S,T) = \sum_{x\in G}\mathbf{1}\big\{|S\cap P_i(x)| = |S\cap P_j(x)| = r,\, S\triangle T = P_{ij}(x)\big\}$$

if $|P_{ij}(0)| = 4r$, and $M_{ij}(S,T) = 0$ if $|P_{ij}(0)| \neq 4r$; note that, despite the asymmetry in their definition, these matrices are in fact symmetric. We will next deduce from inequality (8) a lower bound on the expectation (11).

For a vector $Z \in \{-1,1\}^G$, denote by $Z^{\odot s} \in \{-1,1\}^{\binom{G}{s}}$ the "lifted" vector given by

$$Z^{\odot s}(S) = \prod_{y\in S} Z(y) \quad \text{for all } S \in \binom{G}{s}.$$

If $|P_{ij}(0)| = 4r$, then for all $Z \in \{-1,1\}^G$ we have

$$\begin{aligned}\sum_{S,T\in\binom{G}{s}} M_{ij}(S,T)Z^{\odot s}(S)Z^{\odot s}(T) &= \sum_{S,T\in\binom{G}{s}} M_{ij}(S,T)\prod_{y\in S\triangle T} Z(y)\\ &= \sum_{x\in G}\sum_{S\in\binom{G}{s}} \mathbf{1}\big\{|S\cap P_i(x)| = |S\cap P_j(x)| = r\big\}\prod_{y\in P_{ij}(x)} Z(y)\\ &= \binom{2r}{r}^2\binom{N-4r}{s-2r}\sum_{x\in G}\prod_{y\in P_{ij}(x)} Z(y), \end{aligned} \tag{12}$$

since there are $\binom{2r}{r}^2\binom{N-4r}{s-2r}$ ways of choosing a set $S \in \binom{G}{s}$ satisfying $|S\cap P_i(x)| = |S\cap P_j(x)| = r$ and, once such a set S is chosen, there is only one set $T \in \binom{G}{s}$ for which $S\triangle T = P_{ij}(x)$. It follows that

$$\begin{aligned}&\mathbb{E}_{\substack{\sigma\in\{-1,1\}^L\\ \tau\in\{-1,1\}^R}}\Big\|\sum_{i\in L,j\in R}\sigma_i\tau_j M_{ij}\Big\|_{\infty\to 1}\\ &\qquad\geqslant \mathbb{E}_{\substack{\sigma\in\{-1,1\}^L\\ \tau\in\{-1,1\}^R}}\max_{Z\in\{-1,1\}^G}\sum_{S,T\in\binom{G}{s}}\sum_{i\in L,j\in R}\sigma_i\tau_j M_{ij}(S,T)Z^{\odot s}(S)Z^{\odot s}(T)\\ &\qquad= \mathbb{E}_{\substack{\sigma\in\{-1,1\}^L\\ \tau\in\{-1,1\}^R}}\max_{Z\in\{-1,1\}^G}\binom{2r}{r}^2\binom{N-4r}{s-2r}\sum_{\substack{i\in L,j\in R\\ |P_{ij}(0)|=4r}}\sigma_i\tau_j\sum_{x\in G}\prod_{y\in P_{ij}(x)} Z(y);\end{aligned}$$

combining this with inequalities (8) and (9), we conclude the lower bound

$$\mathbb{E}_{\substack{\sigma\in\{-1,1\}^L\\ \tau\in\{-1,1\}^R}}\Big\|\sum_{i\in L,j\in R}\sigma_i\tau_j M_{ij}\Big\|_{\infty\to 1} \gg_{k,\varepsilon} \binom{N-4r}{s-2r}m^2N. \tag{13}$$

3.4 Applying a Khintchine-type inequality

Now we need to compute an upper bound for the expectation above. The main idea here is to use the non-commutative version of Khintchine's inequality given in Theorem 3. Intuitively, this inequality shows that the sum in the last expression incurs many cancellations due to the presence of the random signs σ_i, and thus the expectation on the left-hand side of (13) is much smaller than one might expect.

To apply Theorem 3, it is better to collect the matrices M_{ij} into groups and use only one half of the random signs σ_i (another idea from [1]). For $i \in L$, $\tau \in \{-1,1\}^R$, we define the matrix

$$M_i^\tau = \sum_{j\in R}\tau_j M_{ij}.$$

We will then provide an upper bound for the expression

$$\max_{\tau\in\{-1,1\}^R}\mathbb{E}_{\sigma\in\{-1,1\}^L}\Big\|\sum_{i\in L}\sigma_i M_i^\tau\Big\|_{\infty\to 1}$$

which is itself an upper bound for the expectation in (13).

Towards this goal, we will prune the matrices M_i^τ by removing all rows and columns whose ℓ_1-weight significantly exceeds the average. By symmetry and non-negativity of these matrices, the ℓ_1-weight of a row or column indexed by a set $S \in \binom{G}{s}$ is bounded by

$$\sum_{T\in\binom{G}{s}} \Big|\sum_{j\in R} \tau_j M_{ij}(S,T)\Big| \leqslant \sum_{T\in\binom{G}{s}} \sum_{j\in R} M_{ij}(S,T)$$
$$= \sum_{\substack{j\in R\\ |P_{ij}(0)|=4r}} \sum_{x\in G} \mathbf{1}\big\{|S\cap P_i(x)| = |S\cap P_j(x)| = r\big\}.$$

To show that pruning makes little difference to the final bounds, we show that only a small proportion of the rows and columns have large ℓ_1-weight. To this end, let U be a uniformly distributed $\binom{G}{s}$-valued random variable and, for each $i \in L$, define the random variable corresponding to the last expression above,

$$X_i := \sum_{\substack{j\in R\\ |P_{ij}(0)|=4r}} \sum_{x\in G} \mathbf{1}\big\{|U\cap P_i(x)| = |U\cap P_j(x)| = r\big\}.$$

The calculation done in (12), with Z the all-ones vector, shows that

$$\mathbb{E}[X_i] = \frac{1}{\binom{N}{s}} \sum_{\substack{j\in R\\ |P_{ij}(0)|=4r}} \binom{2r}{r}^2 \binom{N-4r}{s-2r} N \ll_k \frac{1}{\binom{N}{s}} \binom{N-4r}{s-2r} mN.$$

Since $s = \lfloor N^{1-2/k}\rfloor$, we have that $\binom{N-4r}{s-2r}/\binom{N}{s} \ll_k (s/N)^{2r} \asymp N^{-(2-2/k)}$ and thus

$$\mathbb{E}[X_i] \ll_k \frac{m}{N^{1-2/k}}. \tag{14}$$

The following lemma gives an upper-tail estimate on X_i, provided m is sufficiently large.

Lemma 9. *Suppose that $m \geqslant N^{1-2/k}$. Then, for every $i \in L$, we have that*

$$\Pr\Big[X_i \geqslant (\log N)^k \frac{m}{N^{1-2/k}}\Big] \leqslant \frac{1}{N^4}.$$

Proof. Fix an $i \in L$. Consider the hypergraph H_i on vertex set G and with edge set

$$E(H_i) = \biguplus_{\substack{j\in R\\ |P_{ij}(0)|=4r}} \biguplus_{x\in G} \binom{P_i(x)}{r} \times \binom{P_j(x)}{r},$$

and let $f : \mathbb{R}^G \to \mathbb{R}$ be the polynomial associated with H_i as in (3),

$$f(t) = \sum_{e\in E(H_i)} \prod_{v\in e} t_v.$$

Note that $X_i = f(1_U)$, where U is uniformly distributed over $\binom{G}{s}$ and $1_U \in \mathbb{R}^G$ denotes its (random) indicator vector.

For each $0 \leqslant \ell \leqslant 2r$, we wish to bound the quantity

$$\mu_\ell := \max_{A \in \binom{G}{\ell}} \mathbb{E}_{t \sim \mathrm{Bern}(s/N)^G} f_A(t).$$

(Recall the notation introduced in Section 2.) By (14), we have that $\mu_0 \ll_k mN^{-(1-2/k)}$. For a set $A \in \binom{G}{\ell}$, define its degree in H_i by

$$\deg(A) = |\{e \in E(H_i) : e \supseteq A\}|,$$

where we count multiplicities of repeated edges. Note that for any $B \subseteq A$, we have that $\deg(A) \leqslant \deg(B)$. Then,

$$\mu_\ell = \max_{A \in \binom{G}{\ell}} \left(\frac{s}{N}\right)^{2r-\ell} \deg(A).$$

For any $v \in G$, we have that $\deg(v) \ll_k m$, since v is contained in $O_k(1)$ arithmetic progressions of length k with a fixed common difference. It follows that for $\ell \in [r]$, we have that

$$\mu_\ell \leqslant \left(\frac{s}{N}\right)^{2r-\ell} \max_{v \in G} \deg(v) \ll_k mN^{-2r/(2r+1)} = \frac{m}{N^{1-1/k}}.$$

Let $A \subseteq G$ be a set of size $\ell \in \{r+1, \dots, 2r\}$ and

$$e \in \binom{P_i(x)}{r} \times \binom{P_j(x)}{r}$$

be an edge of $E(H_i)$ that contains A. By the Pigeonhole principle, A contains an element $a \in P_i(x)$ and an element $b \in P_j(x)$. Knowing a limits x to a set of size at most $2r$. Moreover, it follows from (10) that for each x, there are at most $O_k(\log N)$ possible values of $j \in R$ such that $b \in P_j(x)$. Therefore,

$$\mu_\ell \ll_k \left(\frac{s}{N}\right)^{2r-\ell} \log N \leqslant \log N.$$

Using our assumption on m, it follows that for each $\ell \in \{0, \dots, 2r\}$, we have that $\mu_\ell \ll_k mN^{-(1-2/k)} \log N$. The result now follows directly from Corollary 6. □

Lemma 9 shows that for each matrix M_i^τ, at most an N^{-4} fraction of all rows and columns have ℓ_1-weight exceeding $(\log N)^k mN^{-(1-2/k)}$. Now define $\widetilde{M_i^\tau}$ as the 'pruned' matrix obtained from M_i^τ by zeroing out all such heavy rows and columns. Note that $\widetilde{M_i^\tau}$ is symmetric, and so

$$\|\widetilde{M_i^\tau}\|_2 \leqslant \|\widetilde{M_i^\tau}\|_{1\to 1} = \max_{S \in \binom{G}{s}} \|\widetilde{M_i^\tau}(S, \cdot)\|_1 \leqslant (\log N)^k \frac{m}{N^{1-2/k}};$$

this bound on the operator norm is what makes the pruned matrices more convenient for us to work with.

We first show that replacing the original matrices by their pruned versions has negligible effect on our bounds. Indeed, from the definition of X_i we see that its maximum value is bounded by mN, and so

$$\begin{aligned}\big\|M_i^\tau - \widetilde{M}_i^\tau\big\|_{\infty\to 1} &\leqslant \sum_{S\in\binom{G}{s}} \big\|M_i^\tau(S,\cdot) - \widetilde{M}_i^\tau(S,\cdot)\big\|_1 \\ &\leqslant 2\binom{N}{s}\cdot \mathbb{E}\big[X_i\,\mathbf{1}\big\{X_i \geqslant (\log N)^k mN^{-(1-2/k)}\big\}\big] \\ &\leqslant 2\binom{N}{s}\cdot mN\Pr\big[X_i \geqslant (\log N)^k mN^{-(1-2/k)}\big].\end{aligned}$$

(The multiplication by 2 in the second inequality happens because we must take into account both heavy rows and heavy columns.) By Lemma 9 we conclude that

$$\big\|M_i^\tau - \widetilde{M}_i^\tau\big\|_{\infty\to 1} \leqslant \frac{2m}{N^3}\binom{N}{s} \quad \text{for all } i\in L,\ \tau\in\{0,1\}^R. \tag{15}$$

Next we apply the concentration inequality from Theorem 3 to the pruned matrices $\widetilde{M}_i^\tau$; we obtain

$$\begin{aligned}\mathbb{E}_{\sigma\in\{-1,1\}^L}\Big\|\sum_{i\in L}\sigma_i\widetilde{M}_i^\tau\Big\|_{\infty\to 1} &\leqslant \binom{N}{s}\mathbb{E}_{\sigma\in\{-1,1\}^L}\Big\|\sum_{i\in L}\sigma_i\widetilde{M}_i^\tau\Big\|_2 \\ &\leqslant 10\binom{N}{s}\sqrt{\log\binom{N}{s}}\Big(\sum_{i\in L}\|\widetilde{M}_i^\tau\|_2^2\Big)^{1/2} \\ &\leqslant 10\binom{N}{s}\sqrt{\log\binom{N}{s}}\Big(\sum_{i\in L}\|\widetilde{M}_i^\tau\|_{1\to 1}^2\Big)^{1/2} \\ &\leqslant 10\binom{N}{s}\sqrt{s\log N}\cdot m^{1/2}(\log N)^k\frac{m}{N^{1-2/k}}.\end{aligned}$$

By the triangle inequality and our previous bounds, we conclude that

$$\begin{aligned}\mathbb{E}_{\sigma\in\{-1,1\}^L}\Big\|\sum_{i\in L}\sigma_i M_i^\tau\Big\|_{\infty\to 1} &\leqslant \mathbb{E}_{\sigma\in\{-1,1\}^L}\Big\|\sum_{i\in L}\sigma_i\widetilde{M}_i^\tau\Big\|_{\infty\to 1} + \sum_{i\in L}\big\|M_i^\tau - \widetilde{M}_i^\tau\big\|_{\infty\to 1} \\ &\leqslant 10\binom{N}{s}\sqrt{s\log N}\cdot m^{1/2}(\log N)^k\frac{m}{N^{1-2/k}} + \frac{2m^2}{N^3}\binom{N}{s}.\end{aligned}$$

3.5 Finishing the proof

We are now essentially done, and it only remains to combine the upper and lower bounds obtained. Indeed, combining the last inequality with equation (13) gives

$$\binom{N-4r}{s-2r}m^2N \ll_{k,\varepsilon} \binom{N}{s}\sqrt{ms\log N}(\log N)^k\frac{m}{N^{1-2/k}}.$$

Rearranging and using that $\binom{N}{s}/\binom{N-4r}{s-2r} \ll_k (N/s)^{2r} = N^{2-2/k}$, we conclude that

$$m \ll_{k,\varepsilon} s(\log N)^{2k+1} = N^{1-2/k}(\log N)^{2k+1}.$$

As we started with the assumption (5), this shows that $m^*_{k-1,\varepsilon}(G) \ll_{k,\varepsilon} N^{1-2/k}(\log N)^{2k+1}$ as wished.

4 Discussion

Our bounds on $m^*_t(N)$ are far from the conjectured $\Theta_t(\log N)$, and we do not believe that they are best possible. We quickly mention a few avenues that could be explored to obtain better bounds, focusing on the case $m^*_2(N)$ concerning 3-APs for clarity:

- A possible source of inefficiency in our arguments is that, after the symmetrization step (Lemma 7), the fact that $D \in G^m$ is a random sequence is not used in any important way.[2] An improvement on our bound for $m^*_2(N)$ might follow from a possible discrepancy between the worst-case D considered in the present proof and the average-case setting appearing in the problem.

- Another possibility is via a multilinear version of the non-commutative Khintchine inequality to directly bound the final expression in Lemma 7 for a fixed sequence D. Endow the space of trilinear forms (or tensors) $\mathbb{R}^N \times \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ with the norm

  $$\|T\| = \sup\Big\{|T(x,y,z)| \mid \|x\|_3, \|y\|_3, \|z\|_3 \leqslant 1\Big\}.$$

  For trilinear forms $T_1, \dots, T_m$, a bound of the form

  $$\mathbb{E}_{\sigma\in\{-1,1\}^m}\Big\|\sum_{i=1}^m \sigma_i T_i\Big\| \leqslant C(N)\Big(\sum_{i=1}^m \|T_i\|^2\Big)^{\frac{1}{2}}$$

  would imply that $m^*_2(N) \ll C(N)^2$.

  The techniques used in the present paper establish the best bounds currently known for *permutation tensors* of the form

  $$T(x,y,z) = \sum_{i=1}^N x_i y_{\pi_1(i)} z_{\pi_2(i)},$$

  where $\pi_1, \pi_2 \in S_N$ are permutations; this case is sufficient to deal with the forms Λ_D appearing in our proofs. We believe that the bound obtained this way for permutation tensors is not best possible, and a sharper bound for this problem could lead to improvements for $m^*_2(N)$. However, this avenue by itself does not suffice to prove a statement of the form $m^*_2(N) \leqslant \operatorname{poly}\log(N)$, as there is a sequence of permutation tensors (originating from LDC constructions) that imply that necessarily, $C(N) \geqslant (\log N)^{\omega(1)}$ [9, 10].

[2] It is used in a weak way to obtain the technical conditions (9) and (10), but those are mostly technicalities inessential to the main argument.

- The main technical tool used in the present paper is the non-commutative Khintchine inequality (Theorem 3). This inequality is sharp in general, but can be improved when the collection of matrices considered is highly non-commutative; see [22, Section 7] for a discussion on this point. Our matrices are quite close to being commutative, however, and so a possible route for improvement could be to find truly non-commutative matrix embeddings.

Acknowledgements

This work was supported by the Dutch Research Council (NWO) as part of the NETWORKS programme (grant no. 024.002.003). An extended abstract of this work appeared in the proceedings of EUROCOMB'23 [6].

References

[1] Omar Alrabiah, Venkatesan Guruswami, Pravesh K. Kothari, and Peter Manohar. A near-cubic lower bound for 3-query locally decodable codes from semirandom csp refutation. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, STOC 2023, page 1438–1448, New York, NY, U.S.A., 2023. Association for Computing Machinery.

[2] Avraham Ben-Aroya, Oded Regev, and Ronald de Wolf. A hypercontractive inequality for matrix-valued functions with applications to quantum computing and ldcs. In *2008 IEEE 49th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 477–486. IEEE Computer Society, 2008.

[3] V. Bergelson and A. Leibman. Polynomial extensions of van der Waerden's and Szemerédi's theorems. *J. Amer. Math. Soc.*, 9(3):725–753, 1996.

[4] B. Bollobás and A. Thomason. Threshold functions. *Combinatorica*, 7(1):35–38, 1987.

[5] J. Bourgain. Ruzsa's problem on sets of recurrence. *Israel J. Math.*, 59(2):150–166, 1987.

[6] Jop Briët and Davi Castro-Silva. Raising the roof on the threshold for Szemerédi's theorem with random differences. In Daniel Král' and Jaroslav Nešetřil, editors, *Proceedings of the 12th European Conference on Combinatorics, Graph Theory and Applications*, EUROCOMB 2023, pages 231–237. Masaryk University, Brno, 2023.

[7] Jop Briët, Zeev Dvir, and Sivakanth Gopi. Outlaw distributions and locally decodable codes. *Theory Comput.*, 15:Paper No. 12, 24, 2019.

[8] Jop Briët and Sivakanth Gopi. Gaussian width bounds with applications to arithmetic progressions in random settings. *Int. Math. Res. Not. IMRN*, (22):8673–8696, 2020.

[9] Jop Briët, Assaf Naor, and Oded Regev. Locally decodable codes and the failure of cotype for projective tensor products. *Electron. Res. Announc. Math. Sci.*, 19:120–130, 2012.

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

[11] Michael Christ. On random multilinear operator inequalities, 2011. arXiv:1108.5655.

[12] Nikos Frantzikinakis, Emmanuel Lesigne, and Máté Wierdl. Random sequences and pointwise convergence of multiple ergodic averages. *Indiana Univ. Math. J.*, 61(2):585–617, 2012.

[13] Nikos Frantzikinakis, Emmanuel Lesigne, and Máté Wierdl. Random differences in Szemerédi's theorem and related results. *J. Anal. Math.*, 130:91–133, 2016.

[14] Harry Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. Analyse Math.*, 31:204–256, 1977.

[15] W Timothy Gowers. Hypergraph regularity and the multidimensional szemerédi theorem. *Annals of Mathematics*, 166:897–946, 2007.

[16] Iordanis Kerenidis and Ronald de Wolf. Exponential lower bound for 2-query locally decodable codes via a quantum argument. *J. Comput. System Sci.*, 69(3):395–420, 2004. Preliminary version in STOC'03.

[17] Jeong Han Kim and Van H. Vu. Concentration of multivariate polynomials and its applications. *Combinatorica*, 20(3):417–434, 2000.

[18] Brendan Nagle, Vojtěch Rödl, and Mathias Schacht. The counting lemma for regular k-uniform hypergraphs. *Random Structures Algorithms*, 28(2):113–179, 2006.

[19] Vojtěch Rödl and Jozef Skokan. Regularity lemma for k-uniform hypergraphs. *Random Structures Algorithms*, 25(1):1–42, 2004.

[20] E. Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arith.*, 27:199–245, 1975.

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

[22] Joel A. Tropp. An introduction to matrix concentration inequalities. *Foundations and Trends in Machine Learning*, 8(1-2):1–230, 2015.

[23] Trevor D. Wooley and Tamar D. Ziegler. Multiple recurrence and convergence along the primes. *Amer. J. Math.*, 134(6):1705–1732, 2012.