ARITHMETIC EXPANDERS AND DEVIATION BOUNDS FOR SUMS OF RANDOM TENSORS

JOP BRIËT AND SHRAVAS RAO

Abstract. We prove hypergraph variants of the celebrated Alon–Roichman theorem on spectral expansion of sparse random Cayley graphs. One of these variants implies that for every odd prime p and any $\varepsilon > 0$, there exists a set of directions $D \subseteq \mathbb{F}_p^n$ of size $O_{p,\varepsilon}(p^{(1-1/p+o(1))n})$ such that for every set $A \subseteq \mathbb{F}_p^n$ of density α, the fraction of lines in A with direction in D is within $\varepsilon\alpha$ of the fraction of all lines in A. Our proof uses new deviation bounds for sums of independent random multi-linear forms taking values in a generalization of the Birkhoff polytope. The proof of our deviation bound is based on Dudley's integral inequality and a probabilistic construction of ε-nets. Using the polynomial method we prove that a Cayley hypergraph with edges generated by a set D as above requires $|D| \geq \Omega_p(n^{p-1})$ for (our notion of) spectral expansion for hypergraphs.

1. Introduction

In the following all graphs are undirected and may have loops and parallel edges. For an n-vertex graph $G = (V, E)$ and $u, v \in V$ denote by $e_G(u, v)$ the number of edges connecting u and v. If G is k-regular then its normalized adjacency matrix $A_G \in \mathbb{R}^{V \times V}$ is given by $A_G(u, v) = e_G(u, v)/k$. Let $1 = \lambda_1(G) \geq \lambda_2(G) \geq \cdots \geq \lambda_n(G) \geq -1$ be the eigenvalues of A_G arranged in decreasing order and denote $\lambda(G) = \max_{i\in\{2,\ldots,n\}} |\lambda_i(G)|$.

1.1. **Spectral expanders.** Spectral expanders are infinite families of graphs $\{G_i\}_{i\in\mathbb{N}}$ of size increasing with i such that the spectral gap $1 - \lambda(G_i)$ is at least some $\delta > 0$ that is independent of $i \in \mathbb{N}$. A single graph is said to be an expander if it is tacitly understood to belong to such a family. Spectral expansion, the property of having large spectral gap, occurs in random graphs have with high probability. Seminal work on quasirandomness of Thomason [Tho87a, Tho87b], and Chung, Graham, and Wilson [CGW89] showed that for dense graphs, this property is equivalent to a number of other likely features of random graphs. One of these is *expansion*, a measure of connectedness showing that no large set of vertices can be disconnected from its complement by cutting only a few edges. Another is *discrepancy*, which refers the property that the edge density of any sufficiently large induced subgraph is close to the overall edge density.

A long line of research extending the results of [CGW89] to dense hypergraphs was initiated by Chung and Graham [CG90], culminating in recent work of Lenz and Mubayi [LM15b, LM15a] (which we refer to for a more detailed account). Partially motivated by an application in Theoretical Computer Science concerning special types of error-correcting codes (locally decodable codes) [BDG17], we study the extent to which some known results on *sparse*

J. B. was supported by a VENI grant from the Netherlands Organisation for Scientific Research (NWO).
This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1342536.

expanders generalize to hypergraphs. Along the way we establish a new deviation inequality for sums of independent random multi-linear forms (Theorem 2.4) that we hope will find applications elsewhere.

1.2. **Cayley graphs and the Alon–Roichman Theorem.** Most known examples of sparse expanders are Cayley graphs, which are defined as follows. For a finite group Γ and an element $g \in \Gamma$, the Cayley graph $\mathrm{Cay}(\Gamma, \{g\})$ is the 2-regular graph with vertex set Γ and edge set $\{\{u, gu\} : u \in \Gamma\}$, where in case $g^2 = 1$, all edges are doubled. For a multiset[1] $S = \{g_1, \dots, g_k\} \subseteq \Gamma$, the Cayley graph $\mathrm{Cay}(\Gamma, S)$ is the $2k$-regular graph formed by the union of the graphs $\mathrm{Cay}(\Gamma, \{g_1\}), \dots, \mathrm{Cay}(\Gamma, \{g_k\})$.

The group over which Cayley graphs are defined strongly influences the minimal degree required for spectral expansion. The famous examples of constant-degree expanders of Margulis [Mar73, Mar88] and Lubotzky, Phillips, and Sarnak [LPS88] are Cayley graphs which, crucially, are defined over non-Abelian groups. It is easy to see that a Cayley graph over the Abelian group $\mathbb{F}_2^n$, for example, requires degree at least n to be an expander [AR94].

Proposition 1.1. *Let $G = \mathrm{Cay}(\mathbb{F}_2^n, S)$ be such that $|S| < n$. Then, $\lambda(G) = 1$.*

Proof: Let $\Gamma = \mathbb{F}_2^n$. Let $T \subseteq \Gamma$ be an $(n-1)$-dimensional subspace containing S and let $\overline{T} = \Gamma \setminus T$. Since $u, v \in \Gamma$ are connected if and only if $u - v \in S$ and every pair $u \in T, v \in \overline{T}$ satisfies $v - u \in \overline{T}$, the sets T and $\overline{T}$ are disconnected. It follows that $1_\Gamma - 2 1_T$ is an eigenvector of $A(G)$ and has eigenvalue -1. Hence, $\lambda(G) = 1$. □

Similarly, because expanders must be connected, it follows that spectral expansion requires degree $\Omega(\log n)$ in any Cayley graph over any n-element Abelian group [HLW06, Proposition 11.5]. A celebrated result of Alon and Roichman [AR94], however, shows that Abelian groups are extreme in this sense.

Theorem 1.2 (Alon–Roichman Theorem)**.** *For any $\varepsilon \in (0, 1)$ there exists a $c(\varepsilon) \in (0, \infty)$ such that the following holds. Let Γ be a finite group of cardinality n. Let $k \geq c(\varepsilon) \log n$ be an integer and let $g_1, \dots, g_k$ be independent uniformly distributed elements from Γ. Then, with probability at least $1/2$, the Cayley graph $G = \mathrm{Cay}(\Gamma, \{g_1, \dots, g_k\})$ satisfies $\lambda(G) \leq \varepsilon$.*

Our main results are hypergraph versions of Proposition 1.1 and Theorem 1.2.

1.3. **Hypergraphs.** A t-uniform hypergraph $H = (V, E)$ with vertex set V has as edge set E a family of unordered t-element multisets with possible parallel edges. For $u_1, \dots, u_t \in V$ let $e_H(u_1, \dots, u_t)$ denote the number of edges equal to $\{u_1, \dots, u_t\}$. The *adjacency form* of H is the t-linear form $\overline{A}_H : \mathbb{R}^V \times \cdots \times \mathbb{R}^V \to \mathbb{R}$ defined by $\overline{A}_H(1_{\{u_1\}}, \dots, 1_{\{u_t\}}) = e_H(u_1, \dots, u_t)$. The degree of a vertex $v \in V$ is defined by $\overline{A}_H(1_{\{v\}}, 1_V, \dots, 1_V)$ and H is k-regular if every vertex has degree exactly k, in which case its normalized adjacency form is $A_H = \overline{A}_H / k$. Of particular importance here are hypergraphs whose edge set is given by a multiset of the form $\{\pi_1(v), \dots, \pi_t(v)\}$, $v \in V$, where $\pi_1, \dots, \pi_t$ are permutations on V. In this case we set

$$e_H(u_1, \dots, u_t) = \sum_{\sigma} \sum_{v \in V} 1_{\{u_1\}}\big(\pi_{\sigma(1)}(v)\big) \cdots 1_{\{u_t\}}\big(\pi_{\sigma(t)}(v)\big), \tag{1}$$

where σ runs over all permutations of $[t] = \{1, \dots, t\}$, giving a $(t!)$-regular hypergraph.

[1] We use curly brackets to delimit *multisets*: unordered lists that may contain repeated elements.

1.4. **Hypergraph spectral expansion.** To define spectral expansion for hypergraphs we build on the following characterisation of $\lambda(G)$. Recall that the Schatten-∞ norm (or spectral norm) of a matrix A is given by $\|A\|_{S_\infty} = \sup_{x,y\in\mathbb{R}^n\setminus\{\mathbf{0}\}} |x^\mathsf{T} Ay|/\|x\|_{\ell_2}\|y\|_{\ell_2}$. If A is symmetric, then this norm is precisely the maximum absolute value of the eigenvalues of A. Since for an n-vertex graph G, the eigenvector associated with the first eigenvalue $\lambda_1(G) = 1$ is the normalized all-ones vector $\mathbf{1}/\sqrt{n}$, we have $\lambda(G) = \|A_G - J/n\|_{S_\infty}$, where $J = \mathbf{1}\mathbf{1}^\mathsf{T}$ is the all-ones matrix. Our definition of spectral expansion for hypergraphs is based on the following norm on multilinear forms. For a t-linear form A on $\mathbb{R}^n$ and $p \in [1,\infty]$ define

$$\|A\|_{\ell_p,\dots,\ell_p} = \sup\Big\{\frac{A(x[1],\dots,x[t])}{\|x[1]\|_{\ell_p}\cdots\|x[t]\|_{\ell_p}} : x[1],\dots,x[t] \in \mathbb{R}^n \setminus \{\mathbf{0}\}\Big\}.$$

The notion of spectral expansion we shall use is relative to a fixed regular t-uniform hypergraph K. In particular, for a regular t-uniform hypergraph H, we define

$$\lambda_K(H) = \|A_H - A_K\|_{\ell_t,\dots,\ell_t}. \tag{2}$$

For graphs, this parameter coincides with $\lambda(G)$ if K is the complete graph with all loops.

1.5. **Cayley hypergraphs.** A Cayley hypergraph over a finite group Γ is a disjoint union of particular permutation hypergraphs as mentioned in Section 1.3. Let $q \in (\mathbb{Z}\setminus\{0\})^t$ be an integer vector such that no element of Γ has order q_j for every $j \in [t]$. This ensures that for every $g \in \Gamma$, the maps $u \mapsto u^{q_j} g$ are permutations. For $\mathbf{g} = (g[1],\dots,g[t]) \in \Gamma^t$, we define $\mathrm{Cay}^{(t)}(\Gamma, q, \mathbf{g})$ to be the hypergraph as in Section 1.3 based on the permutations $\pi_j(u) = u^{q_j} g[j]$. For a multiset $S = \{\mathbf{g}_1,\dots,\mathbf{g}_k\} \subseteq \Gamma^t$, we let $\mathrm{Cay}^{(t)}(\Gamma, q, S)$ be the $(t!)k$-regular hypergraph given by the union of $\mathrm{Cay}^{(t)}(\Gamma, q, \{\mathbf{g}_i\})$ for $i \in [k]$.

To connect the above definitions, consider a Cayley hypergraph $K = \mathrm{Cay}^{(t)}(\Gamma, q, S)$. For a subset $S' \subseteq S$, let $H = \mathrm{Cay}^{(t)}(\Gamma, q, S')$ be a sub-hypergraph of K and let $\varepsilon = \lambda_K(H)$. Then, for every set $T \subseteq V$ of density $\tau = |T|/|V|$, we have $|(A_H - A_K)(1_T,\dots,1_T)| \le \varepsilon|T|$. Dividing by $|V|$ shows that the fraction of edges that T induces in H is within $\varepsilon\tau$ of the fraction of edges it induces in K.

1.6. **Translation invariant equations.** To motivate the above definitions we focus on a special class of Cayley hypergraphs that arises from systems of translation invariant equations. Such a system can be given in terms of a matrix $C \in \mathbb{Z}^{s\times t}$ and a vector $q \in (\mathbb{Z}\setminus\{0\})^t$ such that $Cq = 0$. For an Abelian group Γ without elements of order q_j for every $j \in [t]$, we then consider the set of solutions in Γ^t to the linear equations defined by C,

$$\mathrm{Sol}(C) = \big\{\mathbf{h} = (h[1],\dots,h[t]) \in \Gamma^t : C\mathbf{h} = 0\big\}.$$

There is a large body of literature on the problem of bounding the maximum size of a set $A \subseteq \Gamma$ such that $\mathrm{Sol}(C) \cap A^t$ contains only trivial solutions. Well-studied examples involving a single equation (where $s = 1$) include Sidon sets [O'B04], where $C = [1,1,-1,-1]$, and sets without 3-term arithmetic progressions (APs), where $C = [1,-2,1]$ (sometimes referred to as cap sets) [O'B11, San11, EG16]. Sets avoiding a general t-variate translation invariant equation were studied in [Ruz93, Blo12, SS14]. Probably the most-studied examples involving more than one equation are t-term APs [Sze90, Gre07, Tao07, O'B11], where

$$C = \begin{bmatrix} 1 & -2 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & -2 & 1 \end{bmatrix} \in \mathbb{Z}^{(t-2)\times t}. \tag{3}$$

Translation invariance refers to the fact that for every $\mathbf{h}$ in $\mathrm{Sol}(C)$ and every $u \in \Gamma$, the tuple $(q_1u+h[1], \dots, q_tu+h[t])$ belongs to $\mathrm{Sol}(C)$ as well. As such, $\mathrm{Sol}(C)$ is a union of cosets of the subgroup $\{(q_1u, \dots, q_tu) : u \in \Gamma\} \subseteq \Gamma^t$. If $S = \{\mathbf{g}_1, \dots, \mathbf{g}_k\}$ is a set of representatives of these cosets, then the edge set of the hypergraph $K = \mathrm{Cay}^{(t)}(\Gamma, q, S)$ is furnished precisely by the (unordered) tuples in $\mathrm{Sol}(S)$, which leads to the following definition.

Definition 1.3 (Arithmetic expander)**.** Let K be the Cayley hypergraph as above. A multiset $S' \subseteq S$ is a $(C, q, \Gamma, \varepsilon)$-*arithmetic expander* if

$$\lambda_K\big(\mathrm{Cay}^{(t)}(\Gamma, q, S')\big) \leq \varepsilon.$$

The preceding discussion shows that an arithmetic expander has the property that for every set $A \subseteq \Gamma$ of density α, the fraction of solutions in A among the cosets represented by S' is within $\varepsilon\alpha$ of the fraction of all solutions in A. For APs, this means the following. The matrix C as in (3) satisfies $Cq = 0$ for $q = \mathbf{1} = (1, \dots, 1)$, from which it follows that $\mathrm{Sol}(C)$ consists of cosets represented by APs through zero, $S = \{\{0, v, 2v, \dots, (t-1)v\} : v \in \Gamma\}$, which correspond to the possible steps that an AP can take. In this case, an arithmetic expander is thus characterized by a small set of steps $D \subseteq \Gamma$ such that the fraction of APs in any set A taking steps from D gives an accurate estimate of the fraction of all APs in A. The AP matrix C also satisfies $Cq = 0$ for $q = (1, 2, \dots, t)$, from which it follows that $\mathrm{Sol}(C)$ consists of the cosets with representatives given by the points through which $(t+1)$-term APs travel, $S = \{\{u, \dots, u\} : u \in \Gamma\}$. In this case, an arithmetic expander thus estimates the fraction of all APs by the fraction of APs travelling through a small fixed set of points.

2. Our results

2.1. Spectral expansion of Cayley hypergraphs. Our first result is an extension of Proposition 1.1 concerning arithmetic expanders for t-APs where t is a prime.

Theorem 2.1. *For every prime p there exist $\varepsilon(p), \delta(p) \in (0, \infty)$ such that the following holds. Let $n \geq p^2$ be an integer, let $\Gamma = \mathbb{F}_p^n$ and let C be as in* (3) *with $t = p$. Then, for any $\varepsilon < \varepsilon(p)$, any $(C, \mathbf{1}, \Gamma, \varepsilon)$-arithmetic expander has size at least $\delta(p)\, n^{p-1}$.*

Our second result is a version of Theorem 1.2, showing for instance that in the AP case, for C as in (3), there exist $(C, q, \mathbb{F}_p^n, \varepsilon)$-arithmetic expanders of size $c(t,\varepsilon)p^{(1-1/t+o(1))n}$ for both options of q, where $c(t, \varepsilon)$ depends on t and ε only.

Theorem 2.2. *For every integer $t \geq 3$ and $\varepsilon \in (0, 1)$ there exists a $c(t, \varepsilon) \in (0, \infty)$ such that the following holds. Let Γ be a finite group of cardinality n, let $q \in (\mathbb{Z} \setminus \{0\})^t$ be such that Γ has no elements of order q_j for every $j \in [t]$, let $S \subseteq \Gamma^t$ be a multiset and $K = \mathrm{Cay}^{(t)}(\Gamma, q, S)$. For $k = c(t, \varepsilon)n^{1-1/t}(\log n)^{t+1/2}$, let $S' \subseteq S$ be a multi-set of k independent uniformly distributed tuples from S. Then, with probability at least $1/2$, $\lambda_K(\mathrm{Cay}^{(t)}(\Gamma, q, S')) \leq \varepsilon$.*

2.2. **A deviation bound for sums of random tensors.** Our proof of Theorem 2.2 follows similar lines as a slick proof of Theorem 1.2 due to Landau and Russel [LR04]. Their proof is based on a matrix-valued deviation inequality called the matrix-Chernoff bound. One can also use the following matrix version of the Hoeffding bound, which follows from a non-commutative Khintchine inequality of Tomczak-Jaegermann [TJ74] (see Appendix A) and which is more in line with the tools we shall use below.

Theorem 2.3 (Matrix Hoeffding bound)**.** *There exist absolute constants $c, C \in (0,\infty)$ such that the following holds. Let $A_1,\dots,A_k \in \mathbb{R}^{n\times n}$ be independent random matrices such that $\|A_i\|_{S_\infty} \le 1$ for each $i \in [k]$. Then, for any $\varepsilon > 0$, we have*

$$\Pr\Big[\Big\|\frac{1}{k}\sum_{i=1}^{k}\big(A_i - \mathbb{E}[A_i]\big)\Big\|_{S_\infty} > \varepsilon\Big] \le C\exp\Big(-\frac{ck\varepsilon^2}{\log n}\Big).$$

Proof of Theorem 1.2: For $s \in \Gamma$ let $A_s \in \mathbb{R}^{\Gamma\times\Gamma}$ denote the adjacency matrix of the Cayley graph $\mathrm{Cay}(\Gamma,\{s\})$. Since A_s is the average of two permutation matrices, $\|A_s\|_{S_\infty} \le 1$. Observe that if $s \in \Gamma$ is a uniformly distributed element, then $\mathbb{E}[A_s] = J/n$. Moreover, since $A_G = (A_{g_1} + \cdots + A_{g_k})/k$, the result now follows from Theorem 2.3. □

The proof of Theorem 2.2 is similarly based on a new deviation bound for multi-linear forms that belong to a generalization of the Birkhoff polytope (of doubly-stochastic matrices). To define this polytope, we first consider the following generalization of a doubly-stochastic matrix. Let $e_1,\dots,e_n \in \mathbb{R}^n$ be the standard basis vectors and let $\mathbf{1} \in \mathbb{R}^n$ denote the all-ones vector. A t-linear form A on $\mathbb{R}^n$ is *plane sub-stochastic* if A is nonnegative on the standard basis vectors and if for every $s \in [n]$, we have

$$\begin{aligned} A(e_s,\mathbf{1},\mathbf{1},\dots,\mathbf{1}) &\le 1\\ A(\mathbf{1},e_s,\mathbf{1},\dots,\mathbf{1}) &\le 1\\ &\vdots\\ A(\mathbf{1},\mathbf{1},\dots,\mathbf{1},e_s) &\le 1. \end{aligned} \tag{4}$$

Let $\Pi_n^{(t)}$ be the polytope of t-linear forms A on $\mathbb{R}^n$ such that the form $|A|$ defined by $|A|(e_{s_1},\dots,e_{s_t}) = |A(e_{s_1},\dots,e_{s_t})|$, for $s_1,\dots,s_t \in [n]$, is plane sub-stochastic. Observe that the set $\Pi_n^{(2)}$ is the set of matrices $(a_{ij})_{i,j=1}^n$ such that $(|a_{ij}|)_{i,j=1}^n$ is doubly sub-stochastic.[2] Our deviation bound then is as follows.

Theorem 2.4. *For every integer $t \ge 3$ there exist absolute constants $c, C \in (0,\infty)$ such that the following holds. Let $A_1,\dots,A_k$ be independent random elements over $\Pi_n^{(t)}$. Then, for any $p \ge 1$ and $\varepsilon > 0$,*

$$\Pr\Big[\Big\|\frac{1}{k}\sum_{i=1}^{k}(A_i - \mathbb{E}[A_i])\Big\|_{\ell_p,\dots,\ell_p} > \varepsilon\Big] \le C\exp\Big(-\frac{ck\varepsilon^2}{\sigma_{p,t}(n)^2}\Big), \tag{5}$$

where

$$\sigma_{p,t}(n) = n^{\frac{1}{2}-\frac{1}{p}}\,\max\{1, n^{1-\frac{1}{2t}-\frac{t-1}{p}}\}\,(\log n)^{t+\frac{1}{2}}.$$

[2]Recall that the Birkhoff–von Neumann Theorem states that the Birkhoff polytope is the convex hull of the set of $n\times n$ permutation matrices. In [LL14] it is shown that for $t \ge 3$, the natural analogue of this *fails* for the set of forms in $\Pi_n^{(t)}$ that attain equalities in (4) and are nonnegative on standard basis vectors.

For example, for $t \geq 3$, we have $\sigma_{2,t}(n) = (\log n)^{t+\frac{1}{2}}$, $\sigma_{t,t}(n) = n^{\frac{1}{2}-\frac{1}{2t}}(\log n)^{t+\frac{1}{2}}$ and $\sigma_{\infty,t}(n) = n^{\frac{3}{2}-\frac{1}{2t}}(\log n)^{t+\frac{1}{2}}$. The proof of Theorem 2.2 is now nearly identical to the proof of the Alon–Roichman theorem shown above.

Proof of Theorem 2.2: Let $H = \mathrm{Cay}^{(t)}(\Gamma, q, S')$. For $\mathbf{g} \in \Gamma^t$ let $A_{\mathbf{g}} : \mathbb{R}^\Gamma \times \cdots \times \mathbb{R}^\Gamma \to \mathbb{R}$ be the adjacency form of $\mathrm{Cay}(\Gamma, q, \{\mathbf{g}\})$ and recall from Section 1.5 that $A_{\mathbf{g}}$ is plane sub-stochastic. Observe that if $\mathbf{g}$ is uniform over S, then $\mathbb{E}[A_{\mathbf{g}}] = A_K$. Finally, since $A_H = (A_{\mathbf{g}_1} + \cdots + A_{\mathbf{g}_k})/k$, the result follows from Theorem 2.4 (with $p = t$) and the definition of $\lambda_K(H)$. $\square$

Remark 2.5 (Sub-optimality of Theorem 2.4?). We conjecture that when $p = t$, the dependence of (5) on n is sub-optimal in the sense that $\sigma_{t,t}$ can be replaced with some function $f(n) \leq o(n^{\frac{1}{2}-\frac{1}{2t}})$. However, due to a result of Naor, Regev, and the first author [BNR12] (see also [Bri15]), it must be the case that $f(n) \geq (\log n)^c$ for every $c > 1$. Their result implies that for every $t \geq 3$ there exist $\varepsilon(t) \in (0, 1)$, $c(t) > 1$ such that the following holds. For infinitely many $n \in \mathbb{N}$, there exists a collection of $k = 2^{(\log\log n)^{c(t)}}$ forms $B_1, \ldots, B_k \in \Pi_n^{(t)}$ such that for independent Rademacher random variables $\epsilon_1, \ldots, \epsilon_k$ (satisfying $\Pr[\epsilon_i = +1] = \Pr[\epsilon_i = -1] = 1/2$), we have

$$\mathbb{E}\Big[\Big\|\frac{1}{k}\sum_{i=1}^{k} \epsilon_i B_i\Big\|_{\ell_t,\ldots,\ell_t}\Big] \geq \varepsilon(t).$$

Setting $A_i = \epsilon_i B_i$, a standard calculation shows that the above expectation is at most

$$\int_0^\infty \Pr\Big[\Big\|\frac{1}{k}\sum_{i=1}^{k}(A_i - \mathbb{E}[A_i])\Big\|_{\ell_t,\ldots,\ell_t} > \varepsilon\Big]\, d\varepsilon \leq C\sqrt{\frac{f(n)}{k}}$$

for some absolute constant $C \in (0, \infty)$, showing that f cannot be poly-logarithmic in n.

Open problems. Our results leave open the problem of determining the minimal degree required for spectral expansion of random Cayley hypergraphs. Remark 2.5 could be interpreted as suggesting the intriguing possibility that, in stark contrast with the Alon–Roichman Theorem, this degree must be quasi-polynomial in the size of the group. Another problem is to determine the optimal form of Theorem 2.4. Finally, it is open if the straightforward generalization of the Expander Mixing Lemma given in Proposition 3.1 below admits a converse for Cayley hypergraphs. A converse was shown to hold for Cayley graphs by Kohayakawa, Rödl, and Schacht [KRS16] and Conlon and Zhao [CZ16].

Acknowledgements. J. B. would like to thank Jozef Skokan for pointing him to [LM15b] and Zeev Dvir and Sivakanth Gopi for helpful discussions.

3. Proof of Theorem 2.1

In this section we prove Theorem 2.1. To rephrase this result, consider for a set $D \subseteq \mathbb{F}_p^n$ the Cayley hypergraph

$$L_D = \mathrm{Cay}^{(p)}\big(\mathbb{F}_p^n, \mathbf{1}, \{\{0, y, 2y, \ldots, (p-1)y\} : y \in D\}\big).$$

Then, by Definition 1.3, Theorem 2.1 says that for every $D \subseteq \mathbb{F}_p^n$ of size $|D| < \delta(p)n^{p-1}$, the hypergraphs L_D and $L_{\mathbb{F}_p^n}$ satisfy $\lambda_{L_{\mathbb{F}_p^n}}(L_D) \geq \varepsilon(p)$. The first ingredient of the proof is the following straightforward generalization of the Expander Mixing Lemma [AC88], which follows directly from the above definitions.

Proposition 3.1 (Generalized Expander Mixing Lemma). *Let $K = \mathrm{Cay}^{(t)}(\Gamma, q, S)$ be a Cayley hypergraph, $S' \subseteq S$ be a multiset and $H = \mathrm{Cay}^{(t)}(\Gamma, q, S')$. Then, for every $T_1, \ldots, T_t \subseteq \Gamma$,*

$$\left|A_H(1_{T_1}, \ldots, 1_{T_t}) - A_K(1_{T_1}, \ldots, 1_{T_t})\right| \leq \lambda_K(H)(|T_t| \cdots |T_t|)^{1/t}.$$

To prove the theorem it thus suffices to show that for every $D \subseteq \mathbb{F}_p^n$ of size $|D| < \delta(p)n^{p-1}$, there exist $T_1, \ldots, T_p \subseteq \mathbb{F}_p^n$ such that on the one hand, $A_{L_D}(1_{T_1}, \ldots, 1_{T_p}) = 0$, while on the other hand, $A_{L_{\mathbb{F}_p^n}}(1_{T_1}, \ldots, 1_{T_p}) \geq \varepsilon(p)(|T_1| \cdots |T_p|)^{1/p}$, which is precisely what we shall do with sets satisfying $T_2 = T_3 = \cdots = T_p$. We achieve this by constructing a combinatorial rectangle $R = T_1 \times \cdots \times T_p$ that contains many lines, but no lines with direction in D, by which we mean the following. Define the *line through $x \in \mathbb{F}_p^n$ in direction $d \in \mathbb{F}_p^n$*, denoted $\ell_{x,d}$, to be the sequence $(x + \lambda d)_{\lambda \in \mathbb{F}_p}$. Say that R *contains* $\ell_{x,d}$ if $x + \lambda d \in T_{\lambda+1}$ for every $\lambda \in \mathbb{F}$. Denote by $\mathcal{L}_D(R)$ the number of lines contained in R that have direction $y \in D$. The following proposition shows why considering lines through rectangles suffices.

Proposition 3.2. *Let $D, T_1, T_2 \subseteq \mathbb{F}_p^n$ so that T_1 and T_2 are disjoint, and let R be the p-dimensional combinatorial rectangle $T_1 \times T_2 \times \cdots \times T_2$. Then,*

$$A_{L_D}(1_{T_1}, 1_{T_2}, \ldots, 1_{T_2}) = \mathcal{L}_D(R)/|D|.$$

Proof: Recall from Section 1.5 and multi-linearity, that

$$\begin{aligned} A_{L_D}(1_{T_1}, 1_{T_2} \ldots, 1_{T_2}) &= \sum_{z_1 \in T_1} \cdots \sum_{z_p \in T_p} A_{L_D}(1_{\{z_1\}}, \ldots, 1_{\{z_p\}}) \\ &= \frac{1}{|D|p!} \sum_{y \in D} \sum_{x \in \mathbb{F}_p^n} \sum_{\sigma \in S_p} 1_{T_1}\big(x + (\sigma(1) - 1)y\big) \cdots 1_{T_2}\big(x + (\sigma(p) - 1)y\big). \end{aligned} \tag{6}$$

Consider a pair $x \in \mathbb{F}_p^n, y \in D$ such that the corresponding sum over $\sigma \in S_p$ in (6) is nonzero. We claim that in this case, the sum equals $(p-1)!$. Indeed, if σ is a permutation such that the corresponding term in the sum equals 1, then since T_1 and T_2 are disjoint, a term corresponding to another permutation σ' is nonzero if and only if $\sigma'(1) = \sigma(1)$. Let $P \subseteq \mathbb{F}_p^n \times D$ be the set of such pairs for which the sum over σ is nonzero. It follows that (6) is equal to $|P|/|D|p$ and the lemma follows if $|P| = p\mathcal{L}_D(R)$.

We compute the size of P. Let $\phi : P \to \mathcal{L}_D(R)$ be the function $\phi((x, y)) = \ell_{x+(\sigma(1)-1)y,y}$ that maps a pair in P to a line in R where σ is an arbitrary permutation that contributes to the corresponding sum in (6). To see that the image of ϕ contains only lines in R, observe that for every pair $(x, y) \in P$, and for $\lambda = \sigma(1) - 1$, we have $x + \lambda y \in T_1$ and $x + \lambda' y \in T_2$ for every $\lambda' \in \mathbb{F}_p \smallsetminus \{\lambda\}$. Moreover, ϕ is surjective, since for each line in $\ell_{x,y} \in \mathcal{L}_D(R)$, we have $(x, y) \in P$ because the term corresponding to the identity permutation in (6) is nonzero.

Next we show that for each $\ell_{x,y} \in \mathcal{L}_D(R)$, its pre-image under ϕ has size exactly p, which implies the proposition. Let (x', y') be a pair in $\phi^{-1}(\ell_{x,y})$. Then $y' = y$ is fixed, and $x' = x + \lambda y$ for some $\lambda \in \mathbb{F}_p$. We claim that all such choices of (x', y') are in $\phi^{-1}(\ell_{x,y})$. Indeed, for every $\lambda \in \mathbb{F}_p$, and $\sigma \in S_p$ such that $\sigma(1) - 1 = \lambda$, we have that $x' + (\sigma(1) - 1)y' \in T_1$ and $x' + (\sigma(i) - 1)y' \in T_2$ for all other i. Therefore $\phi(x', y') = \ell_{x,y}$. □

Theorem 2.1 will thus follow from the following result.

Theorem 3.3. *Let $n \geq p^2$ and let $D \subseteq \mathbb{F}^n$ be a set of size $|D| < \binom{n+p-2}{p-1}$. Then, there exist disjoint sets $T_1, T_2 \subseteq \mathbb{F}^n$ such that the p-dimensional rectangle $T_1 \times T_2 \times \cdots \times T_2$ contains at least p^{2n+p-p^2} lines, but no lines with direction in D.*

Proof of Theorem 2.1: Let $T_1, T_2 \subseteq \mathbb{F}_p^n$ be as in Theorem 3.3. Then, by Proposition 3.2, we have $A_{L_D}(1_{T_1}, 1_{T_2}, \dots, 1_{T_2}) = 0$, but

$$A_{L_{\mathbb{F}_p^n}}(1_{T_1}, 1_{T_2}, \dots, 1_{T_2}) \geq \frac{1}{p^n} p^{2n+p-p^2} \geq \frac{1}{p^{p^2-p}} |T_1|^{1/p} |T_2|^{(p-1)/p}.$$

The result now follows from Proposition 3.1. □

The proof of Theorem 3.3 uses the polynomial method. For the remainder of this section let $\mathbb{F} = \mathbb{F}_p$. For an n-variate polynomial $f \in \mathbb{F}[x_1, \dots, x_n]$ denote $Z(f) = \{x \in \mathbb{F}^n : f(x) = 0\}$.

Lemma 3.4. *Let $f \in \mathbb{F}[x_1, \dots, x_n]$ be a homogeneous polynomial of degree at most $p-1$. Let $Z = Z(f)$ and let $a \in \mathbb{F}^*$ be such that the set $S = \{x \in \mathbb{F}^n : f(x) = a\}$ is nonempty. Then, the p-dimensional rectangle $Z \times S \times \cdots \times S$ contains no lines with directions $d \in Z$.*

Proof: Recall that a Vandermonde matrix is a square matrix of the form

$$\begin{bmatrix} 1 & a_1 & a_1^2 & \cdots & a_1^{d-1} \\ 1 & a_2 & a_2^2 & \cdots & a_2^{d-1} \\ 1 & a_3 & a_3^2 & \cdots & a_3^{d-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & a_d & a_d^2 & \cdots & a_d^{d-1} \end{bmatrix}.$$

We record the well-known and easy fact that if the a_i are distinct, then the above matrix has a nonzero determinant and therefore full rank.

For a contradiction, suppose there does exist such a line $\ell_{x,d}$ with $x, d \in Z$. Consider the polynomial $g \in \mathbb{F}[\lambda]$ defined by $g(\lambda) = a^{-1} f(x + \lambda d)$. Since f has degree at most $p-1$, so does g. Moreover, since $x, d \in Z$ and since f is homogeneous, the constant term and the coefficient of λ^{p-1} of g are zero. Our assumption that $x + \lambda d \in S$ for every $\lambda \in [p-1]$ then implies

$$g(\lambda) = \sum_{i=1}^{p-2} c_i \lambda^i = a^{-1} f(x + \lambda d) = 1, \quad \lambda \in [p-1].$$

Hence, the all-ones vector $\mathbf{1} \in \mathbb{F}^d$ lies in the linear span of the vectors $v_i = (1, 2^i, 3^i, \dots, (p-1)^i)$ for $i \in [p-2]$, since $\mathbf{1} = c_1 v_1 + \cdots + c_{p-1} v_{p-2}$. But the matrix $[\mathbf{1}, v_1, \dots, v_{p-2}]$ is a full-rank Vandermonde matrix, which is a contradiction. □

The following basic and standard result (see for example [Tao14]) shows that for any small set $D \subseteq \mathbb{F}^n$, we can always find a low-degree homogeneous polynomial f such that $D \subseteq Z(f)$.

Lemma 3.5 (Homogeneous Interpolation)**.** *For every $D \subseteq \mathbb{F}^n$ of size $|D| < \binom{n+d-1}{d}$ there exists a nonzero homogeneous polynomial $f \in \mathbb{F}[x_1, \dots, x_n]$ of degree d such that $D \subseteq Z(f)$.*

Proof: Let V be the vector space of homogeneous degree-d polynomials in $\mathbb{F}[x_1, \dots, x_n]$. Note that $\dim(V) = \binom{n+d-1}{n-1}$. Let $W = \mathbb{F}^D$. Let $L : V \to W$ be the linear map given by $L(f) = (f(x))_{x \in D}$. Since $\dim(W) < \dim(V)$, it follows from the Rank Nullity Theorem that $\dim(\ker(L)) \geq 1$. Hence, there exists a nonzero $f \in V$ such that $f(x) = 0$ for all $x \in D$. □

We also use the following standard result bounding the zero-set of a polynomial in terms of its degree; the specific form quoted below is from [CT14, Lemma 2.2].

Lemma 3.6 (DeMillo–Lipton–Schwartz–Zippel)**.** *Let $\mathbb{F}$ be a finite field with q elements and let $f \in \mathbb{F}[x_1, \ldots, x_n]$ be a nonzero polynomial of degree d. Then,*

$$|Z(f)| \leq \Big(1 - \frac{1}{q^{d/(q-1)}}\Big) q^n.$$

Proof: For $g \in \mathbb{F}[x_1, \ldots, x_m]$ write $\overline{Z(g)} = \mathbb{F}^m \setminus Z(g)$. Using induction on n we shall prove that $|\overline{Z(f)}| \geq q^n/q^{d/(q-1)}$, which establishes the result. First observe that, by Lagrange's Theorem, we may assume that each variable in f has degree at most $q-1$. The base case $n = 1$ follows from the Factor Theorem, since then $|\overline{Z(f)}| \geq q - d \geq q(1 - d/q) \geq q/q^{d/(q-1)}$.

Assume the result holds for $(n-1)$-variable polynomials. We can decompose f as

$$f(t, y_1, \ldots, y_{n-1}) = \sum_{i=1}^{\min\{d,q-1\}} t^i g_i(y_1, \ldots, y_{n-1}), \tag{7}$$

where $g_i \in \mathbb{F}[y_1, \ldots, y_{n-1}]$ has degree at most $d - i$. Let k be the maximum i for which g_i is nonzero. By the induction hypothesis, the polynomial g_k satisfies $|\overline{Z(g_k)}| \geq q^{n-1}/q^{(d-k)/(q-1)}$.

For each $y \in \overline{Z(g_k)}$ let $h_y \in \mathbb{F}[t]$ be the univariate polynomial defined by $h_y(t) = f(t, y_1, \ldots, y_{n-1})$. The decomposition (7) shows that each h_y is nonzero and has degree k, and thus $\overline{Z(h_y)} \geq q/q^{k/(q-1)}$. We conclude that

$$|\overline{Z(f)}| \geq \sum_{y \in \overline{Z(g_k)}} |\overline{Z(h_y)}| \geq q^n/q^{d/(q-1)}.$$

□

Finally, we use the Chevalley–Warning Theorem to lower bound the number of common zeros of a system of polynomials [LN83, Chapter 6].

Theorem 3.7 (Chevalley–Warning)**.** *Let $\mathbb{F}$ be a finite field and let $f_1, \ldots, f_k \in \mathbb{F}[x_1, \ldots, x_n]$ be nonzero polynomials such that $d = \deg(f_1) + \cdots + \deg(f_k) < n$. If there is at least one solution to the system $f_1(x) = \cdots = f_k(x) = 0$ in $\mathbb{F}^n$, then there are at least $|\mathbb{F}|^{n-d}$ solutions.*

We include a quick proof we learned from Dion Gijswijt, which is based on Lemma 3.6.

Proof: Define the polynomial $f \in \mathbb{F}[x_1, \ldots, x_n]$ by $f = (1 - f_1^{q-1}) \cdots (1 - f_k^{q-1})$. Observe that $\deg(f) \leq (q-1)d$ and that $f(x) = 1$ if $x \in Z(f_1, \ldots, f_k)$ and $f(x) = 0$ otherwise. By Lemma 3.6, $|Z(f)| \leq (1 - 1/q^d)q^n$. Hence, $|Z(f_1, \ldots, f_k)| \geq q^{n-d}$. □

With this, we are set up to prove Theorem 3.3.

Proof of Theorem 3.3: By Lemma 3.5, there exists a nonzero degree-$(p-1)$ homogeneous polynomial $f \in \mathbb{F}[x_1, \ldots, x_n]$ such that $D \subseteq Z(f)$. Set $T_1 = Z(f)$. By Lemma 3.6, there exists an $a \in \mathbb{F}^*$ such that the set $S = \{x \in \mathbb{F}^n : f(x) = a\}$ is nonempty. For each $\lambda \in [p-1]$ set $T_{\lambda+1} = S$. It then follows from Lemma 3.4 that the combinatorial rectangle $R = T_1 \times \cdots \times T_p$ contains no lines with direction in D.

We show that R contains many lines. To this end, define degree-$(p-1)$ polynomials $g_0, \ldots, g_{p-1} \in \mathbb{F}[x_1, \ldots, x_n, y_1, \ldots, y_n]$ by setting $g_0(x, y) = f(x)$ and $g_\lambda = f(x + \lambda y) - a$ for each $\lambda \in [p-1]$. Then, a solution in $\mathbb{F}^{2n}$ to the set of equations $g_0(x, y) = 0, \ldots, g_{p-1}(x, y) = 0$ is a line through R. There is at least one such solution. Indeed, if we let $x = 0$ and $y \in S$, then since f is homogenous of degree $p-1$, we have $g_0(0, y) = f(0) = 0$ and $g_\lambda(0, y) = f(\lambda y) - a = a(\lambda^{p-1} - 1) = 0$ by Fermat's Little Theorem. By Theorem 3.7, the above system has at least p^{2n+p-p^2} solutions in $\mathbb{F}^{2n}$ and R has at least that many lines. □

4. Proof of Theorem 2.4

In this section we prove Theorem 2.4. Throughout this section, let $(\epsilon_i)_{i\in\mathbb{N}}$ be independent uniformly distributed $\{-1,1\}$-valued random variables and let $\epsilon = (\epsilon_1,\dots,\epsilon_k)$.

4.1. **Reduction to Bernoulli processes.** The main new ingredient needed for the proof of Theorem 2.4 is a bound showing that for fixed $A_1,\dots,A_k \in \Pi_n^{(t)}$, the expected norm of the Rademacher sum $\epsilon_1 A_1 + \cdots + \epsilon_k A_k$ is at most a constant times $\sqrt{k}\,\sigma_{p,t}(n)$. From this, we derive the result using standard techniques based on combining a symmetrization trick, the Kahane–Khintchine inequality and an exponential Markov inequality. The details follow below. Recall that a real-valued random variable is *centered* if it has expectation zero.

The following standard symmetrization lemma allows us to bound the moments of the random variable whose tail we aim to bound in (5) in terms of the moments of the norm of a Rademacher sum of *fixed* plane sub-stochastic forms.

Lemma 4.1 (Symmetrization)**.** *Let X be a real finite-dimensional normed vector space and let $Y_1,\dots,Y_k \subseteq X$ be subsets. For $p \geq 1$ let $\sigma_p \in [0,\infty)$ be the smallest number such that for any fixed $A_1 \in Y_1,\dots,A_k \in Y_k$, we have*

$$\Big(\mathbb{E}\Big[\Big\|\sum_{i=1}^k \epsilon_i A_i\Big\|^p\Big]\Big)^{1/p} \leq \sigma_p.$$

Then, if $A_1,\dots,A_k$ are independent random variables over $Y_1,\dots,Y_k$, respectively,

$$\Big(\mathbb{E}\Big[\Big\|\sum_{i=1}^k (A_i - \mathbb{E}[A_i])\Big\|^p\Big]\Big)^{1/p} \leq 4\sigma_p. \tag{8}$$

Proof: For each $i \in [k]$ let A_i' be an independent copy of A_i. Let $B_i = A_i - \mathbb{E}[A_i']$, let $\widetilde{B}_i$ be an independent copy of B_i and note that these random variables are centered. By Jensen's inequality and the triangle inequality, the pth power of the left-hand side of (8) is at most

$$\mathbb{E}\Big[\Big\|\sum_{i=1}^k B_i\Big\|^p\Big] = \mathbb{E}\Big[\Big\|\sum_{i=1}^k B_i - \mathbb{E}\Big[\sum_{j=1}^k \widetilde{B}_j\Big]\Big\|^p\Big] \leq \mathbb{E}\Big[\Big\|\sum_{i=1}^k (B_i - \widetilde{B}_i)\Big\|^p\Big].$$

Since the random variables $B_i - \widetilde{B}_i$ are independent and symmetrically distributed, that is, $B_i - \widetilde{B}_i$ has the same distribution as $\widetilde{B}_i - B_i$, it follows that their sum has the same distribution as $\delta_1(B_1 - \widetilde{B}_1) + \cdots + \delta_k(B_k - \widetilde{B}_k)$ for any choice of signs $\delta_i \in \{-1,1\}$. Hence, by Jensen's inequality, the above is at most

$$\begin{aligned}\mathbb{E}_\epsilon\Big[\mathbb{E}_{B_i,\widetilde{B}_i}\Big[\Big\|\sum_{i=1}^k \epsilon_i(B_i - \widetilde{B}_i)\Big\|^p\Big]\Big] &\leq 2^p\mathbb{E}_\epsilon\Big[\mathbb{E}_{B_i}\Big[\Big\|\sum_{i=1}^k \epsilon_i B_i\Big\|^p\Big]\Big]\\ &= 2^p\mathbb{E}_\epsilon\Big[\mathbb{E}_{A_i,A_i'}\Big[\Big\|\sum_{i=1}^k \epsilon_i(A_i - \mathbb{E}[A_i'])\Big\|^p\Big]\Big].\end{aligned}$$

Independence of A_i and A'_i for each $i \in [k]$, another application of Jensen's inequality and the triangle inequality imply that the above is at most

$$4^p \mathbb{E}_\epsilon \Big[\mathbb{E}_{A_i} \Big[\Big\| \sum_{i=1}^k \epsilon_i A_i \Big\|^p \Big] \Big] = 4^p \mathbb{E}_{A_i} \Big[\mathbb{E}_\epsilon \Big[\Big\| \sum_{i=1}^k \epsilon_i A_i \Big\|^p \Big] \Big].$$

The result now follows by applying the definition of σ_p to the inner expectation above. $\square$

Next, the Kahane–Khintchine inequality reduces the problem of bounding the numbers σ_p from Lemma 4.1 to bounding σ_1 only (see for example [LT91, Theorem 4.7]).

Theorem 4.2 (Kahane–Khintchine inequality)**.** *Let X be a Banach space and $A_1, \dots, A_k \in X$. Then, for any integer $p \geq 1$ and some absolute constant C, we have*

$$\Big(\mathbb{E} \Big[\Big\| \sum_{i=1}^k \epsilon_i A_i \Big\|^p \Big] \Big)^{1/p} \leq \sqrt{Cp}\, \mathbb{E} \Big[\Big\| \sum_{i=1}^k \epsilon_i A_i \Big\| \Big]. \tag{9}$$

Lemma 4.1 and Theorem 4.2 thus show that the moments on the left-hand side of (8) can be bounded in terms of the average on the right-hand side of (9). The following upper bound and a standard exponential Markov argument will now allow us to prove Theorem 2.4.

Theorem 4.3. *For every integer $t \geq 3$ there exists an absolute constant $C(t) \in (0, \infty)$ such that the following holds. Let $A_1, \dots, A_k \in \Pi_n^{(t)}$. Then, for any $p \geq 1$ and $\sigma_{p,t}(n)$ as in Theorem 2.4,*

$$\mathbb{E} \Big[\Big\| \sum_{i=1}^k \epsilon_i A_i \Big\|_{\ell_p, \dots, \ell_p} \Big] \leq C(t) \sqrt{k}\, \sigma_{p,t}(n) \tag{10}$$

Proof of Theorem 2.4: Define the random variable

$$Z = \Big\| \frac{1}{k} \sum_{i=1}^k (A_i - \mathbb{E}[A_i]) \Big\|_{\ell_p, \dots, \ell_p}.$$

Let $\alpha > 0$ be a parameter to be set later. Then, by Markov's inequality,

$$\Pr[Z > \varepsilon] = \Pr\big[e^{\alpha Z^2} > e^{\alpha \varepsilon^2}\big] \leq e^{-\alpha \varepsilon^2} \mathbb{E}\big[e^{\alpha Z^2}\big]. \tag{11}$$

Lemma 4.1, Theorem 4.2 and Theorem 4.3 imply that for every integer $p \geq 1$, we have

$$\big(\mathbb{E}[Z^p]\big)^{1/p} \leq 4C\sqrt{p}\, C(t) \frac{\sigma_{p,t}(n)}{\sqrt{k}}.$$

Let $\sigma = CC(t)\sigma_{p,t}(n)\sqrt{16/k}$, so that the above equals $\sqrt{p}\,\sigma$. It follows that

$$\begin{aligned} \mathbb{E}\big[e^{\alpha Z^2}\big] &= 1 + \sum_{p=1}^\infty \frac{\alpha^p \mathbb{E}[Z^{2p}]}{p!} \\ &\leq 1 + \sum_{p=1}^\infty \frac{\alpha^p (2p)^p \sigma^{2p}}{p!} \\ &\leq 1 + \sum_{p=1}^\infty \Big(\frac{2\alpha\sigma^2}{e}\Big)^p, \end{aligned}$$

where in the last line we used that $p! \geq (p/e)^p$. Set $\alpha = e/(4\sigma^2)$. Then, the above geometric series equals 2 and the right-hand side of (19) is at most $2e^{-\alpha\varepsilon^2}$, giving the result. $\square$

The remainder of this section is devoted to the proof of Theorem 4.3.

4.2. **Dyadic decomposition.** The first step towards proving Theorem 4.3 is to break the problem up into more manageable pieces using the following lemma. For every $d \in [n]$ define the set

$$\mathcal{C}_d^n = \big\{x \in \{-1,0,1\}^n : \|x\|_{\ell_0} = \min\{d,n\}\big\}. \tag{12}$$

Lemma 4.4. *Let $R = \lceil \log n \rceil$. Then for $p \geq 1$,*

$$\mathbb{E}\Big[\Big\|\sum_{i=1}^k \epsilon_i A_i\Big\|_{\ell_p,\dots,\ell_p}\Big] \leq 2^t \sum_{\mathbf{r}\in[R]^t} \frac{\mathbb{E}\Big[\max\Big\{\Big(\sum_{i=1}^k \epsilon_i A_i\Big)(\mathbf{x}) : \mathbf{x} \in \mathcal{C}_{2^{r_1}}^n \times \cdots \times \mathcal{C}_{2^{r_t}}^n\Big\}\Big]}{2^{\frac{r_1+\cdots+r_t}{p}}}. \tag{13}$$

Proof: Partition the unit ball B_p^n of ℓ_p^n into R pieces defined for each $r \in [R]$ by

$$\mathcal{S}_t(r) = \Big([-\tfrac{2}{2^{r/p}}, -\tfrac{1}{2^{r/p}}) \cup \{0\} \cup (\tfrac{1}{2^{r/p}}, \tfrac{2}{2^{r/p}}]\Big)^n \cap B_p^n \qquad \text{and} \qquad \mathcal{S}_t(R) = [-\tfrac{1}{2^{R/p}}, \tfrac{1}{2^{R/p}}]^n.$$

Then, since each A_i is linear in each of its arguments,

$$\begin{aligned}
\mathbb{E}\Big[\Big\|\sum_{i=1}^k \epsilon_i A_i\Big\|_{\ell_p,\dots,\ell_p}\Big] &= \mathbb{E}\Big[\sup\Big\{\Big(\sum_{i=1}^k \epsilon_i A_i\Big)(\mathbf{x}) : \mathbf{x} \in B_p^n \times \cdots \times B_p^n\Big\}\Big] \\
&\leq \sum_{\mathbf{r}\in[R]^t} \mathbb{E}\Big[\sup\Big\{\Big(\sum_{i=1}^k \epsilon_i A_i\Big)(\mathbf{x}) : \mathbf{x} \in \mathcal{S}_p(r_1) \times \cdots \times \mathcal{S}_p(r_t)\Big\}\Big].
\end{aligned}$$

Note that any $x \in \mathcal{S}_p(r)$ has at most 2^r nonzero entries and that those entries have magnitude at most $2/2^{r/p}$. Hence, by multi-linearity of the A_i and convexity, the above suprema are bounded from above by

$$\frac{2^t}{2^{\frac{r_1+\cdots+r_t}{p}}} \max\Big\{\Big(\sum_{i=1}^k \epsilon_i A_i\Big)(\mathbf{x}) : \mathbf{x} \in \mathcal{C}_{2^{r_1}}^n \times \cdots \times \mathcal{C}_{2^{r_t}}^n\Big\}.$$

$\square$

The above lemma thus reduces the problem of bounding the expectations of Theorem 4.3 to bounding each of the expectations appearing in the right-hand side of (13). The following lemma provides the bounds we need.

Lemma 4.5. *Let $\mathbf{d} \in [n]^t$. Let $\min(\mathbf{d}) = \min\{d_1,\dots,d_t\}$ and let $\max(\mathbf{d}) = \max\{d_1,\dots,d_t\}$. Then, for $\mathcal{C}_d^n$ as in (12), we have*

$$\mathbb{E}\Big[\max\Big\{\Big(\sum_{i=1}^k \epsilon_i A_i(\mathbf{x}) : x[s] \in \mathcal{C}_{d_s}^n,\ s \in [t]\Big\}\Big] \leq C(t)\sqrt{k\max(\mathbf{d})\log n}\ \min(\mathbf{d})^{1-\frac{1}{2t}},$$

where $C(t) \in (0,\infty)$ depends on t only.

Proof of Theorem 4.3: Combining Lemmas 4.4 and 4.5 shows that the left-hand side of (10) is at most

$$C''\sqrt{k\log n} \sum_{r_1,\dots,r_t=1}^{\log n} \frac{\sqrt{\max_s(2^{r_s})}\,\min_s(2^{r_s})^{1-1/(2t)}}{2^{(r_1+\cdots+r_t)/p}}.$$

We claim that each of the above fractions is at most $n^{\frac{1}{2}-\frac{1}{p}}\max\{1, n^{1-\frac{1}{2t}-\frac{t-1}{p}}\}$, from which the claim follows. Indeed, considering the square of these fractions, for any $\mathbf{d} = (d_1, \dots, d_t)$ such that $d_1 \geq d_2 \geq \cdots \geq d_t$, we have

$$\begin{aligned}
\frac{\max(\mathbf{d})\min(\mathbf{d})^{2-1/t}}{(d_1 \cdots d_t)^{2/p}} &= \frac{d_1 d_t^{2-1/t}}{(d_1 \cdots d_t)^{2/p}} \\
&\leq d_1^{1-2/p}\,\frac{d_t^{2-1/t}}{d_t^{2(t-1)/p}} \\
&= d_1^{1-2/p}\max\{1, d_t^{2-1/t-2t/p+2/p}\} \\
&\leq n^{1-2/p}\,\max\{1, n^{2-1/t-2(t-1)/p}\},
\end{aligned}$$

as claimed. □

4.3. **Dudley's integral inequality.** To prove Lemma 4.5 we use Dudley's *integral inequality* (see for example [Tal14, Lemma 2.2.1 and Eq. (2.38)]), which bounds the expected supremum of a stochastic process endowed with a metric space structure in terms of covering numbers. For a metric space (Λ, d) and $\varepsilon > 0$, an *ε-net* is a subset $\Lambda' \subseteq \Lambda$ such that for every $\lambda \in \Lambda$ there exists an $\gamma \in \Lambda'$ with distance $d(\lambda, \gamma) \leq \varepsilon$ and the *covering number* $N(\Lambda, d, \varepsilon)$ is the smallest integer N such that (Λ, d) admits an ε-net of size N. The *diameter* of a metric space (Λ, d) is given by $\operatorname{diam}(\Lambda) = \sup\{d(\lambda, \gamma) : \lambda, \gamma \in \Lambda\}$.

Theorem 4.6 (Dudley's integral inequality)**.** *There exists an absolute constant $C \in (0, \infty)$ such that the following holds. Let Λ be a finite set and $d : \Lambda \times \Lambda \to \mathbb{R}_+$ be a metric on Λ. Let $(X_\lambda)_{\lambda\in\Lambda}$ be a collection of centered random variables such that for every $\lambda, \gamma \in \Lambda$ and any $\varepsilon > 0$, we have*

$$\Pr\big[|X_\lambda - X_\gamma| > \varepsilon\big] \leq 2\exp\Big(-\frac{\varepsilon^2}{d(\lambda,\gamma)^2}\Big).$$

Then,

$$\mathbb{E}\big[\max_{\lambda\in\Lambda} X_\lambda\big] \leq C\int_0^{\operatorname{diam}(\Lambda)} \sqrt{\log N(\Lambda, d, \varepsilon)}\,d\varepsilon.$$

The following set is relevant to our setting:

$$\Lambda_{\mathbf{d}} = \big\{\big(A_1(\mathbf{x}), \dots, A_k(\mathbf{x})\big) : x[s] \in \mathcal{C}^n_{d_s},\ s \in [t]\big\} \subseteq \mathbb{R}^k. \tag{14}$$

For each $\lambda \in \Lambda_{\mathbf{d}}$ consider the (centered) random variable $X_\lambda = \langle \epsilon, \lambda\rangle$, so that the left-hand side of (10) equals $\mathbb{E}[\max_{\lambda\in\Lambda_{\mathbf{d}}} X_\lambda]$. Moreover, for every $\lambda, \gamma \in \Lambda_{\mathbf{d}}$ and $\varepsilon > 0$, we have

$$\Pr\big[|X_\lambda - X_\gamma| > \varepsilon\big] = \Pr\Big[\Big|\sum_{i=1}^k \epsilon_i(\lambda_i - \gamma_i)\Big| > \varepsilon\Big] \leq 2\exp\Big(-\frac{\varepsilon^2}{2\|\lambda-\gamma\|_{\ell_2}^2}\Big), \tag{15}$$

where the second line follows from Hoeffding's inequality [BLM13, Theorem 2.8]. We shall therefore consider the metric space $(\Lambda_{\mathbf{d}}, \ell_2)$. For our setting, the relevant form of Dudley's inequality is then as follows.

Corollary 4.7. *There exists an absolute constant $C \in (0,\infty)$ such that the following holds. Let $\Lambda_{\mathbf{d}} \subseteq \mathbb{R}^k$ be as in* (14)*. Then,*

$$\mathbb{E}\big[\max_{\lambda\in\Lambda_{\mathbf{d}}}\langle \epsilon, \lambda\rangle\big] \leq C\int_0^{\mathrm{diam}(\Lambda_{\mathbf{d}})} \sqrt{\log N(\Lambda_{\mathbf{d}}, \ell_2, \varepsilon)}\, d\varepsilon. \tag{16}$$

Proof: Let d be the metric on $\Lambda_{\mathbf{d}}$ given by $d(\lambda,\gamma) = \sqrt{2}\|\lambda-\gamma\|_{\ell_2}$. Then, the diameter of $\Lambda_{\mathbf{d}}$ under d equals $\sqrt{2}\,\mathrm{diam}(\Lambda_{\mathbf{d}})$. Moreover, since an ε-net for $(\Lambda_{\mathbf{d}}, d)$ is an $\varepsilon/\sqrt{2}$-net for $(\Lambda_{\mathbf{d}}, \ell_2)$ and *vice versa*, we have $N(\Lambda_{\mathbf{d}}, d, \varepsilon) = N(\Lambda_{\mathbf{d}}, \ell_2, \varepsilon/\sqrt{2})$. By (15), Theorem 4.6 and a change of variables, the left-hand side of (16) is therefore at most

$$\begin{aligned} C\int_0^{\sqrt{2}\,\mathrm{diam}(\Lambda_{\mathbf{d}})} \sqrt{\log N(\Lambda_{\mathbf{d}}, d, \varepsilon)}\, d\varepsilon &= C\int_0^{\sqrt{2}\,\mathrm{diam}(\Lambda_{\mathbf{d}})} \sqrt{\log N(\Lambda_{\mathbf{d}}, \ell_2, \varepsilon/\sqrt{2})}\, d\varepsilon \\ &= \sqrt{2}C\int_0^{\mathrm{diam}(\Lambda_{\mathbf{d}})} \sqrt{\log N(\Lambda_{\mathbf{d}}, \ell_2, \delta)}\, d\delta \end{aligned}$$

□

To apply the above result we need a bound on the diameter of $\Lambda_{\mathbf{d}}$ and its covering numbers.

Proposition 4.8. *We have* $\mathrm{diam}(\Lambda_{\mathbf{d}}) \leq \min(\mathbf{d})\sqrt{k}$.

Proof: Because A_i is plane sub-stochastic, $|A_i(\mathbf{x})| \leq \min(\mathbf{d})$ for every $\mathbf{x} \in \mathcal{C}$. Therefore $\Lambda_{\mathbf{d}}$ is a set of k-dimensional vectors whose entries are bounded above in absolute value by $\min(\mathbf{d})$, and the proposition follows. □

4.4. **Bounds on the covering numbers.** To bound the covering numbers of $(\Lambda_{\mathbf{d}}, \ell_2)$ we use a technique akin to Maurey's empirical method for bounding $N(B_{\ell_1^n}, \ell_2, \varepsilon)$, the covering numbers of the unit ball of ℓ_1^n under ℓ_2 (see for instance [CGLP12, Section 1.4]). In particular, for every t-tuple $\mathbf{x}$ as in (14), we use the probabilistic method to show that there exists another t-tuple $\widetilde{\mathbf{x}}$ of vectors $\widetilde{x}[s] \in \mathbb{R}^n$ such that each $\widetilde{x}[s]$ is a "sparse" version of $x[s]$. By this we mean that it has few nonzero entries, each of which has relatively small magnitude, and such that $\mathbf{A}(\widetilde{\mathbf{x}})$ is close to $\mathbf{A}(\mathbf{x})$ in Euclidean distance. This implies that there exists a net composed of all points $\mathbf{A}(\widetilde{\mathbf{x}})$ such that $\widetilde{\mathbf{x}}$ is sparse and that the covering numbers can be bounded by the number of t-tuples of sparse vectors. The sparse vectors themselves are obtained by taking the empirical average of independent samples from a signed and scaled standard basis vector whose average equals $x[s]$. Quantitatively, we get the following lemma.

Lemma 4.9. *For any $\varepsilon \geq 1$, the set $\Lambda_{\mathbf{d}}$ has an $\varepsilon\sqrt{k\min(\mathbf{d})}$-net of size $n^{2t\max(\mathbf{d})/\varepsilon^{2/t}}$.*

Proof: Fix $x[1] \in \mathcal{C}^n_{d_1}, \dots, x[t] \in \mathcal{C}^n_{d_t}$ and let $D_1, \dots, D_t \subseteq [n]$ denote their supports. Let $\eta = \varepsilon^{2/t}$ and do the following for each $s \in [t]$: Set $c_s = d_s/\eta$ (assume for simplicity that this is an integer) and note that $c_s \leq n$. Let $e[s]$ be an independent random standard basis vector of $\mathbb{R}^n$ whose nonzero coordinate is distributed uniformly over D_s. For each $l \in [c_s]$ let $e[s]_l$ be an independent copy of $e[s]$. Define the random vector $\widetilde{x}[s]_l = (x[s] \circ e[s]_l)$, where $\circ$ denotes coordinate-wise multiplication, and define

$$\widetilde{x}[s] = \frac{1}{c_s}\sum_{l=1}^{c_s} d_s\, \widetilde{x}[s]_l.$$

Thus, $\widetilde{x}[s]$ is the empirical average of independent random vectors with expectation $x[s]$. By multi-linearity of the A_i it follows that $\mathbb{E}[\mathbf{A}(\widetilde{\mathbf{x}})] = \mathbf{A}(\mathbf{x})$.

We bound the expected Euclidean distance of $\mathbf{A}(\widetilde{\mathbf{x}})$ and $\mathbf{A}(\mathbf{x})$. By Jensen's inequality,

$$\mathbb{E}\|\mathbf{A}(\widetilde{\mathbf{x}}) - \mathbf{A}(\mathbf{x})\|_{\ell_2} \leq \left(\sum_{i=1}^{k} \mathbb{E}\big[|A_i(\widetilde{\mathbf{x}}) - A_i(\mathbf{x})|^2\big]\right)^{1/2} = \left(\sum_{i=1}^{k} \mathrm{Var}\left[A_i(\widetilde{\mathbf{x}})\right]\right)^{1/2}. \tag{17}$$

We treat each of the above variances separately. Fix an $i \in [k]$. For each $\mathbf{l} \in [c_1] \times \cdots \times [c_t]$ define $\widetilde{\mathbf{x}}_{\mathbf{l}} = \widetilde{x}[1]_{l_1} \times \cdots \times \widetilde{x}[t]_{l_t}$ and $\mathbf{e}_{\mathbf{l}} = e[1]_{l_1} \times \cdots \times e[t]_{l_t}$. Using multi-linearity of A_i, we get

$$\mathbb{E}\big[A_i(\widetilde{\mathbf{x}})^2\big] = \Big(\frac{d_1 \cdots d_t}{c_1 \cdots c_t}\Big)^2 \sum_{\mathbf{l},\mathbf{l}' \in [c]^t} \mathbb{E}\big[A_i(\widetilde{\mathbf{x}}_{\mathbf{l}}) A_i(\widetilde{\mathbf{x}}_{\mathbf{l}'})\big].$$

For every pair $\mathbf{l}, \mathbf{l}' \in [c_1] \times \cdots \times [c_t]$ denote by $\Delta(\mathbf{l}, \mathbf{l}') \subseteq [t]$ the set of coordinates $s \in [t]$ such that $l_s \neq l'_s$. In the case where $|\Delta(\mathbf{l}, \mathbf{l}')| = t$, the tuples $\mathbf{e}_{\mathbf{l}}$ and $\mathbf{e}_{\mathbf{l}'}$ are independent. It follows that in that case, the random variables $A_i(\widetilde{\mathbf{x}}_{\mathbf{l}})$ and $A_i(\widetilde{\mathbf{x}}_{\mathbf{l}'})$ are independent and that

$$\mathbb{E}\big[A_i(\widetilde{\mathbf{x}}_{\mathbf{l}}) A_i(\widetilde{\mathbf{x}}_{\mathbf{l}'})\big] = \frac{A_i(\mathbf{x})^2}{(d_1 \cdots d_t)^2}.$$

Since there are fewer than $(c_1 \cdots c_t)^2$ pairs $\mathbf{l}, \mathbf{l}' \in [c_1] \times \cdots \times [c_t]$ such that $|\Delta(\mathbf{l}, \mathbf{l}')| = t$, the variance is at most

$$\begin{aligned}
\mathrm{Var}\left[A_i(\widetilde{\mathbf{x}})\right] &= \Big(\frac{d_1 \cdots d_t}{c_1 \cdots c_t}\Big)^2 \left(\sum_{|\Delta(\mathbf{l},\mathbf{l}')|<t} \mathbb{E}\big[A_i(\widetilde{\mathbf{x}}_{\mathbf{l}}) A_i(\widetilde{\mathbf{x}}_{\mathbf{l}'})\big] + \frac{1}{(d_1 \cdots d_t)^2} \sum_{|\Delta(\mathbf{l},\mathbf{l}')|=t} A_i(\mathbf{x})^2\right) - A_i(\mathbf{x})^2 \\
&\leq \eta^{2t} \sum_{\Delta(\mathbf{l},\mathbf{l}')<t} \mathbb{E}\big[A_i(\widetilde{\mathbf{x}}_{\mathbf{l}}) A_i(\widetilde{\mathbf{x}}_{\mathbf{l}'})\big].
\end{aligned}$$

Fix a pair $\mathbf{l}, \mathbf{l}' \in [c_1] \times \cdots \times [c_t]$ such that $S = \Delta(\mathbf{l}, \mathbf{l}')$ satisfies $|S| = r < t$ and assume for simplicity that $l_s \neq l'_s$ when $s \leq r$. Since $x[s] \in \mathcal{C}$ for each $s \in [t]$ and since A_i is plane sub-stochastic,

$$\begin{aligned}
\mathbb{E}\big[A_i(\widetilde{\mathbf{x}}_{\mathbf{l}}) A_i(\widetilde{\mathbf{x}}_{\mathbf{l}'})\big] &= \mathbb{E}\left[\left(\prod_{i=1}^{t} \langle x[i], e[i]_{l_i}\rangle \langle x[i], e[i]_{l'_i}\rangle\right) A_i(\mathbf{e}_{\mathbf{l}}) A_i(\mathbf{e}_{\mathbf{l}'})\right] \\
&\leq \mathbb{E}\big[|A_i|(\mathbf{e}_{\mathbf{l}})|A_i|(\mathbf{e}_{\mathbf{l}'})\big] \\
&= \frac{1}{(d_1 \cdots d_r)^2} \mathbb{E}\big[|A_i|(\mathbf{1}_{D_1}, \ldots, \mathbf{1}_{D_r}, e[r+1]_{l_{r+1}}, \ldots, e[t]_{l_t})^2\big] \\
&\leq \frac{1}{(d_1 \cdots d_r)^2} \mathbb{E}\big[|A_i|(\mathbf{1}_{D_1}, \ldots, \mathbf{1}_{D_r}, e[r+1]_{l_{r+1}}, \ldots, e[t]_{l_t})\big] \\
&\leq \frac{|A_i|(\mathbf{1}_{D_1}, \ldots, \mathbf{1}_{D_t})}{(d_1 \cdots d_r)^2 d_{r+1} \cdots d_t} \\
&\leq \frac{\min(\mathbf{d})}{(d_1 \cdots d_r)^2 d_{r+1} \cdots d_t}.
\end{aligned}$$

In general,

$$\mathbb{E}\big[A_i(\widetilde{\mathbf{x}}_{\mathbf{l}}) A_i(\widetilde{\mathbf{x}}_{\mathbf{l}'})\big] \leq \frac{\min(\mathbf{d})}{\prod_{s \in S} d_s^2 \prod_{s \in [t] \setminus S} d_s}.$$

The number of $\mathbf{l}, \mathbf{l}' \in [c_1] \times \cdots \times [c_t]$ such that $\Delta(\mathbf{l}, \mathbf{l}') = S$ is at most $\prod_{s\in S} c_s^2 \prod_{s\in[t]\setminus S} c_s$. Hence, using the definition of c_s and the assumption that $\eta \geq 1$, we get

$$\begin{aligned}
\operatorname{Var}\left[A_i(\widetilde{\mathbf{x}})\right] &\leq \eta^{2t} \sum_{r=0}^{t-1} \sum_{S\in\binom{[t]}{r}} \Big(\prod_{s\in S} c_s^2 \prod_{s\in[t]\setminus S} c_s\Big)\Big(\frac{\min(\mathbf{d})}{\prod_{s\in S} d_s^2 \prod_{s\in[t]\setminus S} d_s}\Big) \\
&\leq \eta^t \min(\mathbf{d}) \sum_{r=0}^{t-1} \binom{t}{r} \eta^{-r} \\
&\leq 2^t \eta^t \min(\mathbf{d}).
\end{aligned}$$

Plugging this into (17), it then follows from the Averaging Principle, there exist vectors $y[s] \in (d_s/c_s)\{-c_s, -c_s+1, \ldots, 0, \ldots, c_s-1, c_s\}$ with at most c_s nonzero entries such that

$$\|\mathbf{A}(y[1], \ldots, y[t]) - \mathbf{A}(x[1], \ldots, x[t])\|_{\ell_2} \leq C_t \eta^{t/2} \sqrt{k \min(\mathbf{d})}.$$

Since there are at most $\prod_{s=1}^t \binom{n}{c_s} c_s^{c_s} \leq n^{2(c_1+\cdots+c_t)} \leq n^{2t\max(\mathbf{d})/\eta}$ tuples $(y[1], \ldots, y[t])$ as above, the result follows. □

4.5. **Putting things together.**

Proof of Lemma 4.5: Lemma 4.9 shows that for any $\varepsilon \geq 1$, we have

$$N\big(\Lambda_{\mathbf{d}}, \ell_2, \varepsilon\sqrt{k\min(\mathbf{d})}\big) \leq n^{2\max(\mathbf{d})/\varepsilon^{2/t}}. \tag{18}$$

In addition, for $\varepsilon > 0$, the size of any $\varepsilon\sqrt{kn}$-net is bounded above by the cardinality of $\Lambda_{\mathbf{d}}$, and therefore

$$N\big(\Lambda_{\mathbf{d}}, \ell_2, \varepsilon\sqrt{k\min(\mathbf{d})}\big) \leq \binom{n}{d_1} \cdots \binom{n}{d_t} \leq n^{2t\max(\mathbf{d})}.$$

Denote $\Delta = \operatorname{diam}(\Lambda_{\mathbf{d}})$. By Corollary 4.7, a substitution of variables, and Proposition 4.8,

$$\begin{aligned}
\int_0^\Delta \sqrt{\frac{\log N(\Lambda_{\mathbf{d}}, \ell_2, \varepsilon)}{2tk\min(\mathbf{d})\max(\mathbf{d})\log n}} d\varepsilon &\leq \int_0^1 d\varepsilon + \int_1^{\Delta/\sqrt{k\min(\mathbf{d})}} \varepsilon^{-1/t} d\varepsilon \\
&\leq 1 + \frac{\Delta^{1-1/t}}{(1-1/t)(k\min(\mathbf{d}))^{1/2-1/(2t)}} \\
&\leq C\min(\mathbf{d})^{1/2-1/(2t)}.
\end{aligned}$$

Hence, there is an absolute constant $C \in (0,\infty)$ such that the left-hand side of (10) is at most

$$C\int_0^\Delta \sqrt{\log N(\Lambda_{\mathbf{d}}, \ell_2, \varepsilon)} d\varepsilon \leq C'\sqrt{k\max(\mathbf{d})\log n}\,\min(\mathbf{d})^{1-1/(2t)}.$$

□

References

[AC88] N. Alon and F. R. Chung. Explicit construction of linear sized tolerant networks. *Annals of Discrete Mathematics*, 38:15–19, 1988.

[AR94] N. Alon and Y. Roichman. Random Cayley graphs and expanders. *Random Structures & Algorithms*, 5, 1994.

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

[BLM13] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities.* Oxford University Press, Oxford, 2013. A nonasymptotic theory of independence, with a foreword by Michel Ledoux.

[Blo12] T. F. Bloom. Translation invariant equations and the method of Sanders. *Bulletin of the London Mathematical Society*, page bds045, 2012.

[BNR12] J. Briët, A. Naor, and O. 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.

[Bri15] J. Briët. On embeddings of ℓ_1^k from locally decodable codes. *Electronic Colloquium on Computational Complexity (ECCC)*, (86), 2015.

[CG90] F. R. K. Chung and R. L. Graham. Quasi-random hypergraphs. *Random Structures & Algorithms*, 1(1):105–124, 1990.

[CGLP12] D. Chafaï, O. Guédon, G. Lecué, and A. Pajor. *Interactions between compressed sensing random matrices and high dimensional geometry*, volume 37 of *Panoramas et Synthèses [Panoramas and Syntheses]*. Société Mathématique de France, Paris, 2012.

[CGW89] F. R. K. Chung, R. L. Graham, and R. M. Wilson. Quasi-random graphs. *Combinatorica*, 9(4):345–362, 1989.

[CT14] G. Cohen and A. Tal. Two structural results for low degree polynomials and applications. *arXiv preprint arXiv:1404.0654*, 2014.

[CZ16] D. Conlon and Y. Zhao. Quasirandom Cayley graphs. *preprint*, 2016. Available at arXiv:1603.03025 [math.CO].

[EG16] J. S. Ellenberg and D. Gijswijt. On large subsets of $\mathbb{F}_q^n$ with no three-term arithmetic progression. *arXiv preprint arXiv:1605.09223*, 2016.

[Gre07] B. Green. Montréal notes on quadratic Fourier analysis. In *Additive combinatorics*, volume 43 of *CRM Proc. Lecture Notes*, pages 69–102. Amer. Math. Soc., Providence, RI, 2007.

[HLW06] S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc.*, 43:439–561, 2006.

[KRS16] Y. Kohayakawa, V. Rödl, and M. Schacht. Discrepancy and eigenvalues of Cayley graphs. *preprint*, 2016. Available at arXiv:1602.02291 [math.CO].

[LL14] N. Linial and Z. Luria. On the vertices of the d-dimensional Birkhoff polytope. *Discrete Comput. Geom.*, 51(1):161–170, 2014.

[LM15a] J. Lenz and D. Mubayi. Eigenvalues and linear quasirandom hypergraphs. In *Forum of Mathematics, Sigma*, volume 3. Cambridge Univ Press, 2015.

[LM15b] J. Lenz and D. Mubayi. The poset of hypergraph quasirandomness. *Random Structures Algorithms*, 46(4):762–800, 2015.

[LN83] R. Lidl and H. Niederreiter. Finite fields. In G.-C. Rota, editor, *Finite Fields*, volume 20 of *Encyclopedia of Mathematics and its Applications*. Addison-Wesley, Reading, Massachusetts, 1983.

[LPS88] A. Lubotzky, R. Phillips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.

[LR04] Z. Landau and A. Russell. Random Cayley graphs are expanders: a simplified proof of the Alon-Roichman theorem. *Electron. J. Combin.*, 11(1), 2004.

[LT91] M. Ledoux and M. Talagrand. *Probability in Banach Spaces.* Springer, 1991.

[Mar73] G. A. Margulis. Explicit constructions of expanders. *Problemy Peredači Informacii*, 9(4):71–80, 1973.

[Mar88] G. A. Margulis. Explicit group-theoretical constructions of combinatorial schemes and their application to the design of expanders and concentrators. *Problems of Information Transmission*, 24(1):39–46, 1988.

[O'B04] K. O'Bryant. A complete annotated bibliography of work related to Sidon sequences. *Electron. J. Combin*, 11:39, 2004.

[O'B11] K. O'Bryant. Sets of integers that do not contain long arithmetic progressions. *Electron. J. Combin.*, 18(1):Paper 59, 15, 2011.

[Ruz93] I. Z. Ruzsa. Solving a linear equation in a set of integers. I. *Acta Arith.*, 65(3):259–282, 1993.

[San11] T. Sanders. On Roth's theorem on progressions. *Ann. of Math. (2)*, 174(1):619–636, 2011.

[SS14] T. Schoen and I. D. Shkredov. Roth's theorem in many variables. *Israel Journal of Mathematics*, 199(1):287–308, 2014.

[Sze90] E. Szemerédi. Integer sets containing no arithmetic progressions. *Acta Math. Hungar.*, 56(1-2):155–158, 1990.

[Tal14] M. 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] T. Tao. The dichotomy between structure and randomness, arithmetic progressions, and the primes. In *International Congress of Mathematicians. Vol. I*, pages 581–608. Eur. Math. Soc., Zürich, 2007.

[Tao14] T. Tao. Algebraic combinatorial geometry: the polynomial method in arithmetic combinatorics, incidence combinatorics, and number theory. *EMS Surv. Math. Sci.*, 1:1–46, 2014.

[Tho87a] A. Thomason. Pseudorandom graphs. In *Random graphs '85 (Poznań, 1985)*, volume 144 of *North-Holland Math. Stud.*, pages 307–331. North-Holland, Amsterdam, 1987.

[Tho87b] A. Thomason. Random graphs, strongly regular graphs and pseudorandom graphs. In *Surveys in combinatorics 1987 (New Cross, 1987)*, volume 123 of *London Math. Soc. Lecture Note Ser.*, pages 173–195. Cambridge Univ. Press, Cambridge, 1987.

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

Appendix A. Proof of Theorem 2.3

For completeness, we derive Theorem 2.3 (the matrix Hoeffding bound) from the following special case of a result of Tomczak-Jaegermann [TJ74, Theorem 3.1].

Theorem A.1 (Tomczak-Jaegermann)**.** *There exists an absolute constant $C \in (0,\infty)$ such that the following holds. Let k be a positive integer and let $p \ge 2$. Let $A_1,\dots,A_k \in \mathbb{R}^{n\times n}$. Then,*

$$\mathbb{E}\Big[\Big\|\sum_{i=1}^k \epsilon_i A_i\Big\|_{S_\infty}\Big] \le C\sqrt{\log n}\Big(\sum_{i=1}^k \|A_i\|_{S_\infty}^2\Big)^{1/2}.$$

Proof of Theorem 2.3: We proceed just as in the proof of Theorem 2.4. For C as in Theorem A.1, let $\sigma = C\sqrt{8(\log n)/k}$ and $\alpha = e/(4\sigma^2)$. Define

$$Z = \Big\|\frac{1}{k}\sum_{i=1}^k (A_i - \mathbb{E}[A_i])\Big\|_{S_\infty}.$$

By Markov's inequality,

$$\Pr[Z > \varepsilon] = \Pr\big[e^{\alpha Z^2} > e^{\alpha\varepsilon^2}\big] \le e^{-\alpha\varepsilon^2}\mathbb{E}\big[e^{\alpha Z^2}\big]. \tag{19}$$

By Lemma 4.1, Theorem 4.2 and Theorem A.1, for every integer $p \ge 1$, $\big(\mathbb{E}[Z^p]\big)^{1/p} \le \sigma\sqrt{p}$. The result then follows since

$$\mathbb{E}\big[e^{\alpha Z^2}\big] \le 1 + \sum_{p=1}^\infty \Big(\frac{2\alpha\sigma^2}{e}\Big)^p = 2.$$

□

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

Courant Institute, New York University, 251 Mercer Street, New York NY 10012, USA
E-mail address: rao@cims.nyu.edu