ON THE BANACH SPACE VALUED AZUMA INEQUALITY AND SMALL-SET ISOPERIMETRY OF ALON–ROICHMAN GRAPHS

ASSAF NAOR

Abstract. We discuss the connection between the expansion of small sets in graphs, and the Schatten norms of their adjacency matrices. In conjunction with a variant of the Azuma inequality for uniformly smooth normed spaces, we deduce improved bounds on the small-set isoperimetry of Abelian Alon–Roichman random Cayley graphs.

1. Introduction

In what follows all graphs are allowed to have multiple edges and loops. For a finite group Γ of cardinality n, the *Cayley graph* associated to the group elements $g_1, \ldots, g_k \in \Gamma$ is the $2k$-regular graph $G = (V, E)$, where $V = \Gamma$, and the number of edges joining $u, v \in \Gamma$ equals $|\{i \in \{1, \ldots, n\} : \ uv^{-1} = g_i\}| + |\{i \in \{1, \ldots, n\} : \ uv^{-1} = g_i^{-1}\}|$.

For a graph $G = (V, E)$ and $S \subseteq V$, let $e_G(S, V \setminus S)$ denote the size of the edge boundary of S, i.e., the number of edges joining S and its complement. The Alon–Roichman theorem [3] asserts that random Cayley graphs obtained by choosing k group elements independently and uniformly at random are good expanders, provided k is large enough:

Theorem 1.1 (Alon–Roichman theorem)**.** *For every $\varepsilon \in (0,1)$ there exists $c(\varepsilon) \in (0, \infty)$ with the following property. Let Γ be a finite group of cardinality n. Assume that $k \geqslant c(\varepsilon) \log n$. Then with probability at least $\frac{1}{2}$ over $g_1, \ldots, g_k$ chosen independently and uniformly at random from Γ, if G is the Cayley graph associated to $g_1, \ldots, g_k$, then for every $\emptyset \neq S \subsetneq \Gamma$ we have*

$$\left| \frac{e_G(S, V \setminus S)}{\frac{2k}{n}|S|(n - |S|)} - 1 \right| \leqslant \varepsilon.$$

Subsequent investigations by several authors [17, 11, 14, 22, 8] yielded new proofs, with various improvements, of the Alon–Roichman theorem. The best known upper bound on $c(\varepsilon)$ is $O(1/\varepsilon^2)$; see [8] for the best known estimate on the implied constant.

Here we obtain an improved estimate on the isoperimetric profile of random Cayley graphs of Abelian groups:

Theorem 1.2. *There exists a universal constant $c \in (0, \infty)$ with the following property. Let Γ be an Abelian group of cardinality n. Assume that $k \geqslant \frac{c \log n}{\varepsilon^2}$. Then with probability at least $\frac{1}{2}$ over $g_1, \ldots, g_k$ chosen independently and uniformly at random from Γ, if G is the Cayley graph associated to $g_1, \ldots, g_k$, then for every $S \subseteq \Gamma$ with $2 \leqslant |S| \leqslant \frac{n}{2}$ we have*

$$\left| \frac{e_G(S, V \setminus S)}{\frac{2k}{n}|S|(n - |S|)} - 1 \right| \leqslant \varepsilon \sqrt{\frac{\log |S|}{\log n}}. \tag{1}$$

Research supported in part by NSF grants CCF-0635078 and CCF-0832795, BSF grant 2006009, and the Packard Foundation.

The possible validity of an estimate such as (1) for finite groups Γ that are not necessarily Abelian remains an interesting open question.

When $k = o(\log n)$, the graphs G of Theorem 1.2 need not be connected, and they are never expanders [3, Prop. 3]. Nevertheless, with positive probability, sufficiently small sets in such graphs do have a large edge boundary:

Theorem 1.3. *There exists a universal constant $c \in (0,\infty)$ with the following property. Let Γ be an Abelian group of cardinality n. Fix $k \in \mathbb{N}$ and $\varepsilon \in (0,1)$. Then with probability at least $\frac{1}{2}$ over $g_1,\ldots,g_k$ chosen independently and uniformly at random from Γ, if G is the Cayley graph associated to $g_1,\ldots,g_k$, then for every $\emptyset \neq S \subsetneq \Gamma$ we have*

$$|S| \leqslant e^{c\varepsilon^2 k} \implies \left| \frac{e_G(S, V \smallsetminus S)}{\frac{2k}{n}|S|(n-|S|)} - 1 \right| \leqslant \varepsilon.$$

The main purpose of this note is *not* to obtain results such as Theorem 1.2 and Theorem 1.3, even if these have independent interest. Our goal is rather to present a method to prove expansion of small sets, going beyond the standard spectral gap techniques. In addition, we highlight a simple and general geometric argument that allows one to reason about such questions for random objects like Alon–Roichman graphs. The rest of this introduction will therefore be devoted to a description these issues. We note that a draft of this manuscript has been circulating for several years, and we were motivated to make it publicly available since the ideas presented here inspired recent progress in theoretical computer science; see [5].

1.1. **Schatten bounds and small-set expansion.** For an n-vertex graph $G = (V,E)$ and $u,v \in V$, let $e(u,v)$ denote the number of edges joining u and v if $u \neq v$, and twice the number of loops at u if $u = v$. If G is d-regular, then the normalized adjacency matrix of G is the $n \times n$ matrix $A(G)$ whose entry at $(u,v) \in V \times V$ equals $e(u,v)/d$. We will denote by $1 = \lambda_1(G) \geqslant \lambda_2(G) \geqslant \cdots \geqslant \lambda_n(G)$ the decreasing rearrangement of the eigenvalues of $A(G)$.

The well-established connection, due to [2, 19] (see also [4, Thm. 9.2.1]), between spectral gaps and graph expansion, reads as follows: for every $\emptyset \neq S \subsetneq V$ we have

$$\left| \frac{e_G(S, V \smallsetminus S)}{\frac{d}{n}|S|(n-|S|)} - 1 \right| \leqslant \max_{i \in \{2,\ldots,n\}} |\lambda_i(G)|. \tag{2}$$

Let $L_2(V)$ denote the vector space $\mathbb{C}^V$, equipped with the standard scalar product

$$\forall\, x,y : V \to \mathbb{C}, \quad \langle x,y \rangle \stackrel{\text{def}}{=} \frac{1}{n} \sum_{u \in V} x(u)\overline{y(u)}.$$

The following lemma is a natural variant of the bound (2).

Lemma 1.4. *Fix $p \in [1,\infty]$. Assume that $L_2(V)$ has an orthonormal basis consisting of eigenvectors of $A(G)$, all of whose entries are bounded by 1 in absolute value. Then for every $\emptyset \neq S \subsetneq V$ we have,*

$$\left| \frac{e_G(S, V \smallsetminus S)}{\frac{d}{n}|S|(n-|S|)} - 1 \right| \leqslant \left(\sum_{i=2}^{n} |\lambda_i(G)|^p \right)^{\frac{1}{p}} \left(\frac{|S|^{\frac{1}{p+1}}(n-|S|)^{\frac{p}{p+1}} + |S|^{\frac{p}{p+1}}(n-|S|)^{\frac{1}{p+1}}}{n} \right)^{\frac{p+1}{p}}.$$

See Lemma 3.1 below for a more general version of Lemma 1.4, which does not require the existence of an orthonormal basis with good L_∞ bounds. We chose to state the above simpler version of Lemma 3.1 in the introduction, since the assumption of Lemma 1.4 holds automatically for Cayley graphs of Abelian groups, where the orthonormal basis in question consists of the characters of the group.

Alon and Roichman [3], as well as the subsequent work [17, 11, 14, 22, 8], proved Theorem 1.1 by showing that, under the assumptions appearing in the statement of Theorem 1.1, we have $\max_{i\in\{2,\ldots,n\}} |\lambda_i(G)| \leqslant \varepsilon$, and then appealing to (2). We will prove Theorem 1.2 and Theorem 1.3 by applying Lemma 1.4 with an appropriate choice of p (depending on the cardinality of S). With this goal in mind, we need to be able to argue about the quantity $\left(\sum_{i=2}^n |\lambda_i(G)|^p\right)^{1/p}$ when G is the random graph appearing in Theorem 1.2. This can be done via simple geometric considerations from Banach space theory.

1.2. **Banach space valued concentration.** The singular values of an $n \times n$ matrix A, i.e., the eigenvalues of $\sqrt{A^*A}$, are denoted $s_1(A) \geqslant s_2(A) \geqslant \cdots \geqslant s_n(A)$. For $p \in [1,\infty]$, the *Schatten p-norm* of A, denoted $\|A\|_{S_p}$, is defined to be $\left(\sum_{i=1}^n s_i(A)^p\right)^{1/p}$. Thus, the quantity $\left(\sum_{i=2}^n |\lambda_i(G)|^p\right)^{1/p}$ appearing in Lemma 1.4 equals $\left\|\left(I - \frac{1}{n}J\right) A(G)\right\|_{S_p}$, where I is the $n \times n$ identity matrix and J is the $n \times n$ matrix all of whose entries are 1.

Fix a group Γ of cardinality n. Let $R : \Gamma \to GL(L_2(\Gamma))$ be the *right regular representation*, i.e., $(R(g)\phi)(h) = \phi(gh)$ for every $\phi : \Gamma \to \mathbb{C}$ and $g, h \in \Gamma$. The normalized adjacency matrix of the Cayley graph associated to $g_1, \ldots, g_k \in \Gamma$ is given by

$$A(g_1, \ldots, g_k) \stackrel{\text{def}}{=} \frac{1}{2k} \sum_{i=1}^k \left(R(g_i) + R(g_i^{-1})\right). \tag{3}$$

In order to apply Lemma 1.4, we are therefore interested in the random quantity

$$\left\| \sum_{i=1}^k \left(I - \frac{1}{n}J\right) \frac{R(g_i) + R(g_i^{-1})}{2k} \right\|_{S_p}. \tag{4}$$

All the known proofs of the Alon–Roichman theorem, corresponding to the case $p = \infty$ in (4), proceed by proving the desired deviation inequality for operator-valued random variables; the original proof of Alon and Roichman uses the Wigner semicircle method, while later proofs rely on the Ahlswede–Winter matrix valued deviation bound [1]. Alternatively, one can use in this context the moment inequalities arising from the non-commutative Khinchine inequalities of Lust-Piquard and Pisier [15, 16], and this method also yields the inequalities that we need for the deviation of the Schatten p-norm in (4). Nevertheless, all of these approaches are specific to operator-valued random variables, and are deeper than the simple argument that we present below. It turns out that for our purposes, it suffices to use an elementary geometric argument that ignores the specific structure of matrix spaces—it works for random variables taking values in arbitrary uniformly smooth normed spaces, of which the Schatten p-norms are a special case.

For a Banach space $(X, \|\cdot\|)$, the triangle inequality implies that $\|x+\tau y\|+\|x-\tau y\| \leqslant 2+2\tau$ for every two unit vectors $x, y \in X$ and every $\tau > 0$. X is said to be *uniformly smooth* if $\|x + \tau y\| + \|x - \tau y\| \leqslant 2 + o(\tau)$, where the $o(\tau)$ term is independent of the choice of unit vectors $x, y \in X$. Formally, consider the following quantity, called the *modulus of uniform*

smoothness of X.

$$\rho_X(\tau) \stackrel{\text{def}}{=} \sup\left\{\frac{\|x+\tau y\|+\|x-\tau y\|}{2}-1:\ x,y\in X,\ \|x\|=\|y\|=1\right\}. \tag{5}$$

Then X is uniformly smooth if $\lim_{\tau\to 0}\frac{\rho_X(\tau)}{\tau}=0$.

X is said to have a modulus of smoothness of power type 2 if there exists $s>0$ such that for all $\tau>0$ we have $\rho_X(\tau)\leqslant s\tau^2$. For simplicity, we will only deal here with spaces that have a modulus of smoothness of power type 2. All of our results below carry over, with obvious modifications, to general uniformly smooth spaces (of course, in this more general setting, the probabilistic bounds that we get will no longer be sub-Gaussian). For concreteness, if $p\geqslant 2$ and S_p denotes the space of all $n\times n$ matrices equipped with the Schatten p-norm, then for every $\tau>0$ we have $\rho_{S_p}(\tau)\leqslant\frac{p-1}{2}\tau^2$. The fact that the modulus of smoothness of S_p has power type 2 when $p\geqslant 2$ was first proved by Tomczak-Jaegermann in [20]. The exact modulus of smoothness of S_p was computed in [6]. The case of $L_p(\mu)$ spaces is much older—the modulus of smoothness in this case was computed by Hanner in [10].

An Azuma-type deviation inequality holds for general norms whose modulus of smoothness has power type 2.

Theorem 1.5. *There exists a universal constant $c\in(0,\infty)$ with the following property. Fix $s>0$ and assume that a Banach space $(X,\|\cdot\|)$ satisfies $\rho_X(\tau)\leqslant s\tau^2$ for all $\tau>0$. Fix also a sequence of positive numbers $\{a_k\}_{k=1}^{\infty}\subseteq(0,\infty)$. Let $\{M_k\}_{k=1}^{\infty}\subseteq X$ be an X-valued martingale satisfying the pointwise bound $\|M_{k+1}-M_k\|\leqslant a_k$ for all $k\in\mathbb{N}$. Then for every $u>0$ and $k\in\mathbb{N}$ we have*

$$\mathbb{P}\left[\|M_{k+1}-M_1\|\geqslant u\right]\leqslant e^{s+2}\cdot\exp\left(-\frac{cu^2}{a_1^2+\cdots+a_k^2}\right). \tag{6}$$

Theorem 1.5 is a consequence of well understood moment inequalities in Banach space theory. The key insights here are due to the work of Pisier. Theorem 1.5 relies on an estimate of implicit constants appearing in Pisier's inequality [18]; this is done in Section 2. While these bounds are not available in [18], undoubtedly Pisier could have computed them if they were needed for the purpose of his investigations in [18]. Therefore, Section 2 should be viewed as mainly expository. In addition to obtaining the estimates that we need, another purpose of Section 2 is to present the proof in a way which highlights the clarity and simplicity of the general geometric argument leading to Theorem 1.5.

Note that the exponential dependence on s in (6) cannot be improved. A roundabout way to see this is to note that when X is the space of $n\times n$ matrices equipped with the Schatten p norm, then since in this case $s\leqslant p/2$, and for $p=\log n$ we have $\|\cdot\|_{S_p}\asymp\|\cdot\|_{S_\infty}$ (= the operator norm), the inequality (6) corresponds (for this value of p) to the Ahlswede–Winter deviation inequality used in [11, 8, 22] to prove the (sharp) logarithmic dependence of k on n in the Alon–Roichman theorem. For random variables taking values in the space of matrices equipped with the operator norm, deeper methods lead to results that are more refined than Theorem 1.5, but do not have an interpretation in the setting of general uniformly smooth Banach spaces, and are not needed for our purposes; see [21] for more information on this topic.

2. The Azuma inequality in uniformly smooth normed spaces

Our main result is the following theorem.

Theorem 2.1. *Fix $s > 0$ and $q \geqslant 2$. Assume that a Banach space X satisfies $\rho_X(\tau) \leqslant s\tau^2$ for all $\tau > 0$. Let $\{Z_n\}_{n=1}^{\infty}$ be X-valued random variables such that for all $n \in N$ we have $\mathbb{E}\left[\|Z_n\|^q\right] < \infty$. Denote $S_n = Z_1 + \cdots + Z_n$. Assume that for all $n \in \mathbb{N}$ we have*

$$\mathbb{E}\left[\|S_n - Z_{n+1}\|^q\right] \geqslant \mathbb{E}\left[\|S_n\|^q\right]. \tag{7}$$

Then

$$\left(\mathbb{E}\left[\|S_n\|^q\right]\right)^{1/q} \leqslant 8\sqrt{s+q} \cdot \sqrt{\sum_{j=1}^{n} \left(\mathbb{E}\left[\|Z_j\|^q\right]\right)^{2/q}}. \tag{8}$$

Before proving Theorem 2.1 we record two concrete examples. The most important case is when $\{Z_n\}_{n=1}^{\infty}$ form a martingale difference sequence. In other words, there exists a filtration $\mathscr{F}_1 \subseteq \mathscr{F}_2 \subseteq \cdots$ such that $Z_1, \ldots, Z_n$ are measurable with respect to $\mathscr{F}_n$ for all $n \in \mathbb{N}$, and for $m > n$ we have $\mathbb{E}\left[Z_m \middle| \mathscr{F}_n\right] = 0$. In this case, using the notation of Theorem 2.1 and the convexity of the norm, we see that

$$\mathbb{E}\left[\|S_n - Z_{n+1}\|^q \middle| \mathscr{F}_n\right] \geqslant \left\|\mathbb{E}\left[S_n - Z_{n+1} \middle| \mathscr{F}_n\right]\right\|^q = \|S_n\|^q. \tag{9}$$

By taking expectation we get that the assumption (7) is satisfied.

Another example worth mentioning is when q is an even integer and $Z_n \in L_q(\Omega, \mu)$ satisfy the point-wise condition $\mathbb{E}\left[Z_{n+1} S_n^{q-1}\right] \leqslant 0$ for all $n \in \mathbb{N}$. In this case

$$\begin{aligned}\mathbb{E}\left[\|S_n - Z_{n+1}\|_{L_q(\Omega,\mu)}^q\right] &= \mathbb{E}\left[\int_\Omega (S_n - Z_{n+1})^q d\mu\right] \\ &\geqslant \mathbb{E}\left[\int_\Omega \left(S_n^q - q Z_{n+1} S_n^{q-1}\right) d\mu\right] \geqslant \mathbb{E}\left[\|S_n\|_{L_q(\Omega,\mu)}^q\right].\end{aligned}$$

Therefore the assumption (7) holds in this case.

Proof of Theorem 1.5. An application of Theorem 2.1 to $Z_n = M_{n+1} - M_n$, together with Markov's inequality, shows that for all $q \geqslant \max\{2, s\}$ and $u > 0$,

$$\mathbb{P}\left[\|M_{k+1} - M_1\| \geqslant u\right] \leqslant \left(\frac{16\sqrt{q(a_1^2 + \cdots + a_k^2)}}{u}\right)^q. \tag{10}$$

The optimal choice of q in (10) is $q = \frac{u^2}{256e(a_1^2 + \cdots + a_k^2)}$, which is an allowed value of q provided $u^2 \geqslant \max\{s, 2\} \cdot 256e(a_1^2 + \cdots + a_k^2)$. This implies (6). □

We now pass to the proof of Theorem 2.1. We start with the following lemma, whose proof is a slight variant of the proof of Proposition 7 in [6].

Lemma 2.2. *Assume that $\rho_X(\tau) \leqslant s\tau^2$ for all $\tau > 0$. Then for every $x, y \in X$ and for every $q \geqslant 2$ we have*

$$\frac{\|x+y\|^q + \|x-y\|^q}{2} \leqslant \left(\|x\|^2 + 8(s+q) \cdot \|y\|^2\right)^{q/2}. \tag{11}$$

Proof. Assume first of all that $\|x\| = 1$ and $\|y\| \leqslant 1$. Denote

$$b \stackrel{\text{def}}{=} \frac{\|x+y\| + \|x-y\|}{2} \quad \text{and} \quad \beta \stackrel{\text{def}}{=} \frac{\|x+y\| - \|x-y\|}{\|x+y\| + \|x-y\|}.$$

Then $b \geqslant 1$, since the function $\tau \mapsto \|x+\tau y\| + \|x-\tau y\|$ is convex and even, and hence attains its minimum at $\tau = 0$. Also, the triangle inequality implies that $\beta \leqslant \frac{\|y\|}{b} \leqslant \|y\| \leqslant 1$. If we write

$$\left(\frac{(1+\beta)^q + (1-\beta)^q}{2}\right)^{1/q} = 1 + \theta$$

then $\theta \in [0,1]$ and $\theta \leqslant (q-1)\beta^2$. Both of these inequalities are elementary numerical facts; the proof of the latter inequality can be found in many places, e.g., [13, Lem. 1.e.14].

Now,

$$\left(\frac{\|x+y\|^q + \|x-y\|^q}{2}\right)^{1/q} = b\left(\frac{(1+\beta)^q + (1-\beta)^q}{2}\right)^{1/q} \stackrel{(\clubsuit)}{\leqslant} [1+\rho_X(\|y\|)]\cdot[1+\theta]$$
$$\stackrel{(\clubsuit\clubsuit)}{\leqslant} \sqrt{1+8\left[\rho_X(\|y\|)+\theta\right]} \leqslant \sqrt{1+8(s+q)\|y\|^2},$$

where in $(\clubsuit)$ we used the definition of ρ_X and θ and in $(\clubsuit\clubsuit)$ we used the elementary inequality $(1+u)(1+v) \leqslant \sqrt{1+8(u+v)}$, which is valid for all $u, v \in [0,1]$. This proves the assertion of Lemma 2.2 when $\|y\| \leqslant \|x\|$. When $\|y\| \geqslant \|x\|$ apply the same reasoning with the roles of x and y reversed, and use the bound $\|y\|^2 + 8(s+q)\cdot\|x\|^2 \leqslant \|x\|^2 + 8(s+q)\cdot\|y\|^2$. $\square$

In what follows, if (Ω, μ) is a measure space then $L_q(X, \Omega, \mu)$ will denote the Banach space of all functions $f : \Omega \to X$ such that

$$\|f\|^q_{L_q(X,\Omega,\mu)} \stackrel{\text{def}}{=} \int_\Omega \|f\|^q d\mu < \infty.$$

In [9] it is shown that if $q \geqslant 2$ and X is 2-smooth then $L_q(X, \Omega, \mu)$ is also 2-smooth. The dependence of the modulus of smoothness of $L_q(X, \Omega, \mu)$ on q and the modulus of smoothness of X can be deduced from the proofs in [9], but is not stated there explicitly. This dependence is crucial for us, so we will now show how it easily follows from Lemma 2.2.

Corollary 2.3. *Assume that $\rho_X(\tau) \leqslant s\tau^2$ for all $\tau > 0$. Then $\rho_{L_q(X,\Omega,\mu)}(\tau) \leqslant 4(s+q)\tau^2$ for every $\tau > 0$.*

Proof. Fix $f, g \in L_q(X, \Omega, \mu)$ with $\|f\|_{L_q(X,\Omega,\mu)} = \|g\|_{L_q(X,\Omega,\mu)} = 1$ and $\tau > 0$. Then,

$$\begin{aligned}
\left(\frac{\|f+\tau g\|_{L_q(X,\Omega,\mu)} + \|f-\tau g\|_{L_q(X,\Omega,\mu)}}{2}\right)^q &\leqslant \frac{\|f+\tau g\|^q_{L_q(X,\Omega,\mu)} + \|f-\tau g\|^q_{L_q(X,\Omega,\mu)}}{2} \\
&= \int_\Omega \frac{\|f+\tau g\|^q + \|f-\tau g\|^q}{2} d\mu \\
&\stackrel{(*)}{\leqslant} \int_\Omega \left(\|f\|^2 + 8(s+q)\cdot\tau^2\|g\|^2\right)^{q/2} d\mu \\
&\stackrel{(**)}{\leqslant} \left(\|f\|^2_{L_q(X,\Omega,\mu)} + 8(s+q)\tau^2\|g\|^2_{L_q(X,\Omega,\mu)}\right)^{q/2} \\
&= \left(1 + 8(s+q)\tau^2\right)^{q/2}, \qquad (12)
\end{aligned}$$

where in $(*)$ we used Lemma 2.2 and in $(**)$ we used the triangle inequality in $L_{q/2}(\Omega,\mu)$. It follows from (12) that

$$\frac{\|f+\tau g\|_{L_q(X,\Omega,\mu)}+\|f-\tau g\|_{L_q(X,\Omega,\mu)}}{2} \leqslant \sqrt{1+8(s+q)\tau^2} \leqslant 1+4(s+q)\tau^2. \qquad \square$$

The following lemma goes back to [12, 7] (see also Proposition 2.2 in [18]).

Lemma 2.4. *Assume that $\rho_X(\tau) \leqslant s\tau^2$ for all $\tau > 0$. Let $\{x_n\}_{n=1}^{\infty} \subseteq X$ be a sequence of vectors in X and for every $n \in \mathbb{N}$ denote $S_n = x_1 + \cdots + x_n$. Assume that for all $n \in \mathbb{N}$ we have $\|S_n - x_{n+1}\| \geqslant \|S_n\|$. Then for every $n \in \mathbb{N}$,*

$$\|S_n\|^2 \leqslant 10(s+2)\left(\|x_1\|^2 + \cdots + \|x_n\|^2\right).$$

Proof. Apply Lemma 2.2 with $q = 2$, $x = S_n$ and $y = x_{n+1}$ to get that

$$\frac{\|S_{n+1}\|^2 + \|S_n\|^2}{2} \overset{(\star)}{\leqslant} \frac{\|S_n + x_{n+1}\|^2 + \|S_n - x_{n+1}\|^2}{2} \leqslant \|S_n\|^2 + 5(s+2)\|x_{n+1}\|^2, \qquad (13)$$

where in $(\star)$ we used the assumption $\|S_n - x_{n+1}\| \geqslant \|S_n\|$. Inequality (13) is equivalent to $\|S_{n+1}\|^2 \leqslant \|S_n\|^2 + 10(s+2)\|x_{n+1}\|^2$. Therefore Lemma 2.4 follows by induction. $\square$

Proof of Theorem 2.1. Let (Ω,μ) be the probability space on which the random variables $\{Z_n\}_{n=1}^{\infty}$ are defined. Lemma 2.4 applied to the Banach space $X' = L_q(X,\Omega,\mu)$, combined with Corollary 2.3, implies Theorem 2.1. $\square$

Remark 2.1. Readers who are mainly interested in the case of operator valued random variables should note that the above argument proves the general Azuma inequality (6) in a self-contained way, except that we quoted the fact that $\rho_{S_p}(\tau) \lesssim p\tau^2$. We wish to stress that the proof of this fact is elementary and accessible to non-experts. When $p = 2k$ is an even integer, the sharp estimate on the modulus of smoothness of S_p was proved in [20] by expanding the quantity $\|A+B\|_{S_p}^p + \|A-B\|_{S_p}^p = \mathrm{trace}\left([(A+B)(A^*+B^*)]^k + [(A-B)(A^*-B^*)]^k\right)$, and estimating the summands in the resulting sum separately. When $p \geqslant 2$ is not an even integer, the computation of the sharp modulus of smoothness of S_p in [6] is more subtle, but still elementary. Note that for the purpose of our application to small-set expansion of graphs, the case of even p suffices. Also, in the case of S_p, the above proof is much shorter (yielding better constants), since the intermediate steps of Lemma 2.2 and Corollary 2.3 are not needed—the inequalities obtained in Lemma 2.2 and Corollary 2.3 are the way that $\rho_{S_p}(\tau)$ was estimated in [20, 6] in the first place. The role of these elementary intermediate steps is only to relate the standard definition (5) of the modulus of uniform smoothness to inequalities such as (11), but in the literature, when one estimates $\rho_X(\tau)$, this is often done by proving (11) directly.

3. Schatten norm bounds and graph expansion

Let $G = (V,E)$ be an n-vertex d-regular graph. For $p \geqslant 1$, the normed space $L_p(V)$ is the space of all $x \in \mathbb{C}^V$, equipped with the norm

$$\|x\|_{L_p(V)} \overset{\mathrm{def}}{=} \left(\frac{1}{n}\sum_{u\in V}|x(u)|^p\right)^{1/p}.$$

In what follows, whenever we refer to an orthornormal eigenbasis $e_1, \ldots, e_n$ of the normalized adjacency matrix $A(G)$, it will always be understood that it is orthonormal in $L_2(V)$, the eigenvectors are indexed so that $A(G)e_j = \lambda_j(G)e_j$ for all $j \in \{1, \ldots, n\}$, and $e_1 = \mathbf{1}_V$ (the constant 1 function).

The following lemma becomes Lemma 1.4 when $q = r = \infty$.

Lemma 3.1. *Assume that $p, r > 1$ and $q > \frac{2p}{p-1}$. Let $G = (V, E)$ be an n-vertex d-regular graph, and let $e_1, \ldots, e_n$ be an orthonormal eigenbasis of the normalized adjacency matrix $A(G)$. Then for every $\emptyset \neq S \subsetneq V$ we have,*

$$\left| \frac{e_G(S, V \setminus S)}{\frac{d}{n}|S|(n-|S|)} - 1 \right| \leqslant M^{\frac{q}{p(q-2)}} \left(\sum_{i=2}^{n} |\lambda_i(G)|^p \right)^{\frac{1}{p}} \left(\frac{|S|^\delta (n-|S|)^{1-\delta} + |S|^{1-\delta}(n-|S|)^\delta}{n} \right)^{\frac{1}{1-\delta}},$$

where

$$M \stackrel{\text{def}}{=} \left(\sum_{j=1}^{n} \|e_j\|_{L_r(V)}^q \right)^{1/q} \quad \text{and} \quad \delta \stackrel{\text{def}}{=} \frac{q(r-2)}{pr(q-2) + q(r-2)}. \tag{14}$$

Proof. Consider the linear operator $T : \mathbb{C}^V \to \mathbb{C}^n$ given by

$$T(x) \stackrel{\text{def}}{=} \Big(\langle x, e_1 \rangle, \langle x, e_2 \rangle, \ldots, \langle x, e_n \rangle \Big).$$

Since $\{e_1, \ldots, e_n\}$ is an orthonormal basis of $L_2(V)$ we have $\|T(x)\|_{\ell_2^n} = \|x\|_{L_2(V)}$ for all $x \in \mathbb{C}^V$. Moreover, using Hölder's inequality we have

$$\|T(x)\|_{\ell_q^n} = \left(\sum_{j=1}^{n} |\langle x, e_j \rangle|^q \right)^{1/q} \leqslant \left(\sum_{j=1}^{n} \|e_j\|_{L_r(V)}^q \|x\|_{L_{\frac{r}{r-1}}(V)}^q \right)^{1/q} = M\|x\|_{L_{\frac{r}{r-1}}(V)}.$$

In other words, we have the operator norm bounds

$$\|T\|_{L_2(V) \to \ell_2^n} = 1 \quad \text{and} \quad \|T\|_{L_{\frac{r}{r-1}}(V) \to \ell_q^n} \leqslant M.$$

Recall that $\frac{2p}{p-1} \in (2, q)$, so we can define $\varepsilon \in (0, 1)$ by $\frac{p-1}{2p} = \frac{\varepsilon}{2} + \frac{1-\varepsilon}{q}$, i.e.,

$$\varepsilon = 1 - \frac{q}{p(q-2)}.$$

If we then define $a > 1$ via $\frac{1}{a} = \frac{\varepsilon}{2} + \frac{(1-\varepsilon)(r-1)}{r}$, i.e.,

$$a = \frac{2pr(q-2)}{pr(q-2) + q(r-2)} \stackrel{(14)}{=} 2(1-\delta),$$

then the Riesz–Thorin interpolation theorem (see [23, Ch. XII]) asserts that for every $x \in \mathbb{C}^V$ we have

$$\|T(x)\|_{\ell^n_{\frac{2p}{p-1}}} \leqslant M^{1-\varepsilon}\|x\|_{L_a(V)} = M^{\frac{q}{p(q-2)}}\|x\|_{L_a(V)}. \tag{15}$$

Fix $S \subseteq V$ and consider the function $x \in \mathbb{C}^V$ given by

$$x \stackrel{\text{def}}{=} (n - |S|)\mathbf{1}_S - |S|\mathbf{1}_{V \setminus S}.$$

Since $\langle x, e_1\rangle = \sum_{u\in V} x(u) = 0$, the bound (15) becomes

$$\left(\sum_{j=2}^{n} |\langle x, e_j\rangle|^{\frac{2p}{p-1}}\right)^{\frac{p-1}{2p}} \leqslant M^{\frac{q}{p(q-2)}} \left(\frac{|S|(n-|S|)^a + (n-|S|)|S|^a}{n}\right)^{\frac{1}{a}}.$$

Hence,

$$\begin{aligned}|\langle A(G)x, x\rangle| = \left|\sum_{j=2}^{n} \lambda_j(G)\langle x, e_j\rangle^2\right| &\leqslant \left(\sum_{j=2}^{n} |\lambda_j(G)|^p\right)^{\frac{1}{p}} \left(\sum_{j=2}^{n} |\langle x, e_j\rangle|^{\frac{2p}{p-1}}\right)^{\frac{p-1}{p}} \\ &\leqslant M^{\frac{2q}{p(q-2)}} \left(\sum_{j=2}^{n} |\lambda_j(G)|^p\right)^{\frac{1}{p}} \left(\frac{|S|(n-|S|)^a + (n-|S|)|S|^a}{n}\right)^{\frac{2}{a}}.\end{aligned}$$

The required result now follows from the identity

$$\langle A(G)x, x\rangle = |S|(n-|S|) - \frac{n}{d} e_G(S, V\setminus S). \qquad \square$$

4. Proof of Theorem 1.2

Let Γ be a group of cardinality n. Recall that $R : \Gamma \to GL(L_2(\Gamma))$ is the right regular representation, and given $g_1, \ldots, g_k \in \Gamma$, the matrix $A(g_1, \ldots, g_k)$ is defined as in (3).

Lemma 4.1. *There exists a universal constant $C \in (0, \infty)$ with the following property. Fix $k, n \in \mathbb{N}$ and a group Γ of cardinality n. Let $g_1, \ldots, g_k \in \Gamma$ be chosen independently and uniformly at random. Then with probability at least $\frac{1}{2}$, for every integer $p \geqslant 2$ we have*

$$\left\|\left(I - \frac{1}{n}J\right) A(g_1, \ldots, g_k)\right\|_{S_p} \leqslant Cn^{1/p}\sqrt{\frac{p}{k}}.$$

Proof. For all $i \in \{1, \ldots, k\}$, since $\mathbb{E}\left[R(g_i)\right] = \mathbb{E}\left[R(g_i^{-1})\right] = \frac{J}{n}$, we have

$$\mathbb{E}\left[\left(I - \frac{J}{n}\right)\frac{R(g_i) + R(g_i^{-1})}{2}\right] = 0. \tag{16}$$

Moreover, note that we have $\left\|\left(I - \frac{1}{n}J\right) R(g_i)\right\|_{S_p} = (n-1)^{1/p}$, because $MM^* = I - \frac{1}{n}J$, where $M = \left(I - \frac{1}{n}J\right) R(g_i)$. We therefore have the (point-wise) bound

$$\left\|\left(I - \frac{1}{n}J\right)\frac{R(g_i) + R(g_i^{-1})}{2}\right\|_{S_p} \leqslant (n-1)^{1/p}. \tag{17}$$

Theorem 2.1 now implies that for some universal constant $c \in (0, \infty)$,

$$\mathbb{E}\left[\left\|\left(I - \frac{1}{n}J\right) A(g_1, \ldots, g_k)\right\|_{S_p}^p\right] \leqslant \left(\frac{cp}{k}\right)^{p/2} n.$$

Hence,

$$\mathbb{E}\left[\sum_{p=2}^{\infty}\left(\frac{\left\|\left(I - \frac{1}{n}J\right) A(g_1, \ldots, g_k)\right\|_{S_p}}{2n^{1/p}\sqrt{cp/k}}\right)^p\right] \leqslant \sum_{p=2}^{\infty}\frac{1}{2^p} = \frac{1}{2}.$$

It follows from Markov's inequality that with probability at least $\frac{1}{2}$ we have

$$\max_{p\in\mathbb{N}\cap[2,\infty)}\left(\frac{\left\|\left(I-\frac{1}{n}J\right)A(g_1,\ldots,g_k)\right\|_{S_p}}{2n^{1/p}\sqrt{cp/k}}\right)^p\leqslant\sum_{p=2}^{\infty}\left(\frac{\left\|\left(I-\frac{1}{n}J\right)A(g_1,\ldots,g_k)\right\|_{S_p}}{2n^{1/p}\sqrt{cp/k}}\right)^p\leqslant 1.\quad\square$$

Proof of Theorem 1.2. Γ is now Abelian, and therefore for every $g_1,\ldots,g_k\in\Gamma$, the characters of Γ are an orthonormal eigenbasis of $A(g_1,\ldots,g_k)$, consisting of functions whose absolute value is point-wise bounded by 1. By Lemma 3.1 and Lemma 4.1, with probability at least $\frac{1}{2}$ over i.i.d. uniform choice of $g_1,\ldots,g_k\in\Gamma$, if we let G be the Alon–Roichman graph whose adjacency matrix is $A(g_1,\ldots,g_k)$, then every $S\subseteq\Gamma$ with $2\leqslant|S|\leqslant\frac{n}{2}$ satisfies

$$\left|\frac{e_G(S,V\smallsetminus S)}{\frac{2k}{n}|S|(n-|S|)}-1\right|\lesssim\min_{p\in\mathbb{N}\cap[2,\infty)}\left(n^{\frac{1}{p}}\sqrt{\frac{p}{k}}\left(\frac{|S|^{\frac{1}{p+1}}(n-|S|)^{\frac{p}{p+1}}+|S|^{\frac{p}{p+1}}(n-|S|)^{\frac{1}{p+1}}}{n}\right)^{\frac{p+1}{p}}\right)$$

$$\lesssim\min_{p\in\mathbb{N}\cap[2,\infty)}\left(|S|^{\frac{1}{p}}\sqrt{\frac{p}{k}}\right)\lesssim\sqrt{\frac{\log|S|}{k}},\quad(18)$$

where in the last step of (18) we choose $p=2\lceil\log|S|\rceil$. $\square$

5. Proof of Theorem 1.3

If $k\leqslant 1/\varepsilon^2$ then Theorem 1.3 is vacuous for $c>0$ small enough. Assuming $k\geqslant 1/\varepsilon^2$, denote $p=2\varepsilon^2k\geqslant 2$. Arguing analogously to (18), we see that there exists a universal constant $K\in(0,\infty)$ such that with probability at least $\frac{1}{2}$ over i.i.d. uniform choice of $g_1,\ldots,g_k\in\Gamma$, if we let G be the Alon–Roichman graph whose adjacency matrix is $A(g_1,\ldots,g_k)$, then for all $S\subseteq\Gamma$ with $2\leqslant|S|\leqslant\frac{n}{2}$,

$$\left|\frac{e_G(S,V\smallsetminus S)}{\frac{2k}{n}|S|(n-|S|)}-1\right|\leqslant K|S|^{\frac{1}{p}}\sqrt{\frac{p}{k}}=\varepsilon K\sqrt{2}|S|^{\frac{1}{2\varepsilon^2k}}\leqslant 2K\varepsilon,$$

provided $|S|\leqslant 2^{\varepsilon^2k}$. $\square$

References

[1] R. Ahlswede and A. Winter. Strong converse for identification via quantum channels. *IEEE Trans. Inform. Theory*, 48(3):569–579, 2002.

[2] N. Alon and V. D. Milman. λ_1, isoperimetric inequalities for graphs, and superconcentrators. *J. Combin. Theory Ser. B*, 38(1):73–88, 1985.

[3] N. Alon and Y. Roichman. Random Cayley graphs and expanders. *Random Structures Algorithms*, 5(2):271–284, 1994.

[4] N. Alon and J. H. Spencer. *The probabilistic method*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience [John Wiley & Sons], New York, second edition, 2000. With an appendix on the life and work of Paul Erdős.

[5] S. Arora, B. Barak, and D. Steurer. Subexponential algorithms for Unique Games and related problems. To appear in proceedings of 51th Annual IEEE Symposium on Foundations of Computer Science, 2010.

[6] K. Ball, E. A. Carlen, and E. H. Lieb. Sharp uniform convexity and smoothness inequalities for trace norms. *Invent. Math.*, 115(3):463–482, 1994.

[7] Buĭ-Min-Či and V. I. Gurariĭ. Certain characteristics of normed spaces and their application to the generalization of Parseval's equality to Banach spaces. *Teor. Funkciĭ Funkcional. Anal. i Priložen. Vyp.*, 8:74–91, 1969.

[8] D. Christofides and K. Markström. Expansion properties of random Cayley graphs and vertex transitive graphs via matrix martingales. *Random Structures Algorithms*, 32(1):88–100, 2008.
[9] T. Figiel. On the moduli of convexity and smoothness. *Studia Math.*, 56(2):121–155, 1976.
[10] O. Hanner. On the uniform convexity of L^p and l^p. *Ark. Mat.*, 3:239–244, 1956.
[11] Z. Landau and A. Russell. Random Cayley graphs are expanders: a simple proof of the Alon-Roichman theorem. *Electron. J. Combin.*, 11(1):Research Paper 62, 6 pp. (electronic), 2004.
[12] J. Lindenstrauss. On the modulus of smoothness and divergent series in Banach spaces. *Michigan Math. J.*, 10:241–252, 1963.
[13] J. Lindenstrauss and L. Tzafriri. *Classical Banach spaces. II*, volume 97 of *Ergebnisse der Mathematik und ihrer Grenzgebiete [Results in Mathematics and Related Areas]*. Springer-Verlag, Berlin, 1979. Function spaces.
[14] P.-S. Loh and L. J. Schulman. Improved expansion of random Cayley graphs. *Discrete Math. Theor. Comput. Sci.*, 6(2):523–528 (electronic), 2004.
[15] F. Lust-Piquard. Inégalités de Khintchine dans C_p $(1 < p < \infty)$. *C. R. Acad. Sci. Paris Sér. I Math.*, 303(7):289–292, 1986.
[16] F. Lust-Piquard and G. Pisier. Noncommutative Khintchine and Paley inequalities. *Ark. Mat.*, 29(2):241–260, 1991.
[17] I. Pak. Random Cayley graphs with $O(\log |G|)$ generators are expanders. In *Algorithms—ESA '99 (Prague)*, volume 1643 of *Lecture Notes in Comput. Sci.*, pages 521–526. Springer, Berlin, 1999.
[18] G. Pisier. Martingales with values in uniformly convex spaces. *Israel J. Math.*, 20(3-4):326–350, 1975.
[19] R. M. Tanner. Explicit concentrators from generalized N-gons. *SIAM J. Algebraic Discrete Methods*, 5(3):287–293, 1984.
[20] N. 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.
[21] J. Tropp. User-friendly tail bounds for matrix martingales. Preprint available at `http://arxiv.org/abs/1004.4389`, 2010.
[22] A. Wigderson and D. Xiao. Derandomizing the Ahlswede-Winter matrix-valued Chernoff bound using pessimistic estimators, and applications. *Theory Comput.*, 4:53–76, 2008.
[23] A. Zygmund. *Trigonometric series. Vol. I, II.* Cambridge Mathematical Library. Cambridge University Press, Cambridge, third edition, 2002. With a foreword by Robert A. Fefferman.

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