Stochastic Construction of Expander Graphs

Po-Shen Loh
Leonard Schulman

October 1, 2003

Abstract

Expander graphs form a class of combinatorial objects that are used for many important constructions that are of interest in the theory of computation; their widespread applications range from error-correcting codes to pseudorandom number generators and switching networks. Yet until recently, constructions of the expander graphs themselves were either highly nontrivial or entirely random and unstructured. In their 2002 paper, Alon and Roichman proved that for every group G of order n, the Cayley graph with respect to $O(\log n)$ random elements almost always produces an expander. This stochastic construction provided a nice alternative to the previously known explicit, but sophisticated, constructions. It also avoided inefficiencies due to local expansion deficiencies in the entirely randomized argument.

However, their proof was fairly complex and derived its bound indirectly, through the use of an auxiliary inequality that related the second-largest (in absolute value) eigenvalue of the graph's adjacency matrix to the probability that a random walk on the graph is closed. In this paper, we present a direct proof of the Abelian case of their result, using only elementary group representation theory (which, incidentally, is equivalent to Fourier analysis in our Abelian case). Furthermore, our representation-theoretic approach induces the stronger conjecture that only $O(\log R)$ generators are required, where R is the number of irreducible representations of G. This would yield an asymptotic improvement because there exist groups, such as the symmetric group, for which R is subpolynomial with respect to n.

1 Background

1.1 Definitions

Definition 1 *Let Γ be an undirected graph with vertex set V and edge set E. We call Γ a* ***c-expander*** *if for any subset $S \subseteq V$, we have $|A(S)| > c|S|(1 - |S|/|V|)$, where $A(S)$ is the set of vertices in $V \setminus S$ that are adjacent to vertices in S.*

Definition 2 *Let G be a group, and let S be a* ***symmetric*** *set of elements (i.e. $s \in S \Leftrightarrow s^{-1} \in S$). The* ***Cayley graph*** *$X(G, S)$ is defined to be the graph with vertex set G and edge set $\{(g, sg) : g \in G, s \in S\}$.*

Definition 3 *Let Γ be a d-regular graph. Then the* ***normalized adjacency matrix*** *of Γ is defined to be $(1/d)\mathbf{A}$, where $\mathbf{A}$ is the adjacency matrix of Γ.*

1.2 Notation

- Let Γ be a graph on n vertices. Its normalized adjacency matrix must be real and symmetric, hence diagonalizable with real eigenvalues; let its eigenvalues counting multiplicity be $(\lambda_1, \lambda_2, \ldots, \lambda_n)$, arranged in order of nonincreasing absolute value. Let $\mu(\Gamma)$ denote $|\lambda_2|$.
- Let $\mathbf{M}$ be a square matrix; let $\lambda(\mathbf{M}) = \max\{|\lambda| : \lambda \text{ is an eigenvalue of } \mathbf{M}\}$.

1.3 Previous work

Several constructions for expander graphs are known, and one important family of such constructions consists of Cayley graphs that turn out to be expanders. These are convenient constructions because the resulting expander graphs inherit additional structure from their parent Cayley graphs. For many years, the only known Cayley graph constructions (e.g. the LPS construction in [6]) explicitly specified groups and generators, but they were very complex and required advanced mathematics.

Then in 2002, Alon and Roichman published (in [1]) a stochastic construction that was as simple as the explicit constructions were intricate; they proved that for any $1 > c > 0$, the Cayley graph of an arbitrary group G with respect to $O(\log |G|)$ random generators was a c-expander, with probability tending to 1 as $|G| \to \infty$. This remarkable expansion result was derived from eigenvalue analysis of the corresponding adjacency matrices, for which they only needed to prove this theorem:

Theorem 1 *(Alon/Roichman, 2002) For every $1 > \delta > 0$, there exists a $C(\delta) > 0$ such that the following holds. Let G be a group, and let S be a set of $C(\delta) \log |G|$ random elements of G. Then*

$$E(\mu(X(G, S))) < 1 - \delta,$$

where the expectation is taken with G fixed and S variable.

The first step of their proof was to use the following general matrix inequality to bound μ, which ignored the fact that the matrices in question had special properties (they were adjacency matrices of Cayley graphs):

Fact 1 *Let A be a real symmetric matrix, with eigenvalues $|\mu_0| \geq |\mu_1| \geq \cdots \geq |\mu_{n-1}|$. Then for every natural number m:*

$$|\mu_1| \leq \left(Tr(A^{2m}) - \mu_0 \right)^{1/2m}$$

Our paper will study Theorem 1 through the lens of group representation theory, with the aim of reproving and refining it. Since our approach will allow us to work with μ directly, it has the potential to recapture information that may have been lost in the application of Fact 1.

2 Our result

2.1 The main theorem

Let G be a finite group, and partition the elements of G into disjoint blocks by pairing the elements with their inverses (and if an element is its own inverse, make it a block by itself). That is, if the elements of G are $\{s_1, s_2, \ldots, s_a, t_1, t_1^{-1}, t_2, t_2^{-1}, \ldots, t_b, t_b^{-1}\}$, where the $s_k = s_k^{-1}$ and the $t_k \neq t_k^{-1}$, then we will obtain the partition:

$$\mathcal{P}(G) = \left\{ \{s_1\}; \{s_2\}; \ldots; \{s_a\}; \{t_1, t_1^{-1}\}; \{t_2, t_2^{-1}\}; \ldots; \{t_b, t_b^{-1}\} \right\}. \tag{1}$$

Observe that if we sample N distinct blocks from $\mathcal{P}(G)$, then we will obtain a symmetric set of elements, with cardinality between N and $2N$ inclusive.

Theorem 2 *Let G be a finite **Abelian** group. Form $\mathcal{P}(G)$ as above, and sample N distinct blocks according to the uniform distribution on the $(a + b)$ blocks. Let S be the union of the N samples. Then for every $\epsilon, \delta \geq 0$, as long as $N \geq C_0(\epsilon, \delta) \log |G|$:*

$$\Pr\left[\mu(X(G, S)) \geq \delta\right] \leq \epsilon, \tag{2}$$

where the function $C_0(\epsilon, \delta)$ depends only on ϵ and δ, and the probability is taken with G fixed and S ranging over all N-block samples from $\mathcal{P}(G)$.

In our proof of this theorem, we will not use the fact that G is Abelian until the very end; this will put us into position to propose the following conjecture regarding the non-Abelian case:

Conjecture 1 *Let G be an **arbitrary** finite group, and let R be its number of distinct irreducible representations. Form $\mathcal{P}(G)$ as above, and sample N distinct blocks according to the uniform distribution on the $(a+b)$ blocks. Let S be the union of the N samples. Then for every $\epsilon, \delta \geq 0$, as long as $N \geq C(\epsilon, \delta) \log R$:*

$$\Pr\left[\mu(X(G,S)) \geq \delta\right] \leq \epsilon, \tag{3}$$

where the function $C(\epsilon, \delta)$ depends only on ϵ and δ, and the probability is taken with G fixed and S ranging over all N-block samples from $\mathcal{P}(G)$.

2.2 Decomposition into irreducible representations

To prove Theorem 2, we begin by passing to irreducible representations. From this point on, let us disregard the fact that G is Abelian. First, consider the group algebra $\mathbb{C}[G]$, and let α be the element in $\mathbb{C}[G]$ defined by:

$$\alpha = \frac{1}{|S|} \sum_{s \in S} s$$

Let the operator T represent the left-action of α on $\mathbb{C}[G]$. If we think of $\mathbb{C}[G]$ as a vector space with basis indexed by G, then the matrix representation of T with respect to this basis coincides the normalized adjacency matrix of the Cayley graph $X(G,S)$. This yields the immediate consequence that T is Hermitian, and shows that it suffices to characterize the second-largest eigenvalue (with respect to absolute value) of T.

Next, we use the representation-theoretic Fourier Transform $\mathcal{F}$, which is an algebra isomorphism from $\mathbb{C}[G]$ to a product of matrix algebras:

$$\mathcal{F} : \mathbb{C}[G] \longrightarrow \prod_{i=1}^{R} \mathcal{M}_i,$$

where R is the number of irreducible representations of G and the $\mathcal{M}_i = \mathrm{Mat}_{d_i \times d_i}(\mathbb{C})$, with the d_i corresponding to the degrees of the various irreducible representations. The explicit formula for $\mathcal{F}$ is:

$$\mathcal{F}\left(\sum_{x \in G} a_x x\right) = \bigoplus_{i=1}^{R} \left(\sum_{x \in G} a_x \rho_i(x)\right),$$

where $\rho_i : G \to \mathcal{M}_i$ are the various irreducible representations, expressed with respect to fixed bases. Without loss of generality, we can choose the bases such that all $\rho_i(x)$ are unitary. Since $\mathcal{F}$ is an isomorphism, the eigenvalues of T are the same as the eigenvalues of the left-action of $\mathcal{F}(\alpha)$ on $\prod \mathcal{M}_i$. Yet by virtue of the direct product, we can now analyze each matrix algebra independently.

Let us focus on an arbitrary component of $\mathcal{F}(\alpha)$, say the one for ρ_i: let $\Psi_i = (1/|S|) \sum_{s \in S} \rho_i(s)$. Note that if λ is an eigenvalue of the action of Ψ_i on $\mathcal{M}_i$, then λ must also be an eigenvalue of the matrix Ψ_i itself. Also note that since Ψ_i is an average of unitary matrices, the absolute values of its eigenvalues must be bounded by 1.

Without loss of generality, suppose that ρ_1 corresponds to the trivial representation $\rho_1 : G \mapsto 1$. In this case, $\Psi_1 = 1$, so it has eigenvalue 1. Therefore, by the previous paragraph,

$$\mu(X(G,S)) = \max_{i \neq 1} \{\lambda(\Psi_i)\} \tag{4}$$

2.3 Analysis of a single irreducible representation

To simplify notation, let us now drop the indices "i." Suppose that we have a nontrivial unitary irreducible representation $\rho : G \to \mathrm{Mat}_{D \times D}(\mathbb{C})$, and let $\Psi = (1/|S|) \sum_{s \in S} \rho(s)$. Since T was Hermitian, so is Ψ; therefore,

$$\lambda(\Psi) = \sup_{\|v\|=1} |\langle \Psi v, v \rangle| . \tag{5}$$

Lemma 1 *There exists a constant c such that for any dimension n and any $\zeta \geq 0$, the **real** unit sphere $S^n \subset \mathbb{R}^{n+1}$ can be covered by an ζ-net of cardinality bounded by $(c/\zeta)^n$. We define a ζ-net to be a collection of points for which every point on S^n is within distance ζ of some point of the net. Note that this means that the **complex** unit sphere $\mathbb{S}^{2n-1} \subset \mathbb{C}^n$ can be covered by a ζ-net of cardinality bounded by $(c/\zeta)^{2n-1}$.*

Proof: We inductively prove this for $c = 2\pi$. For S^1, we clearly can accomplish this with $\lceil \pi/\zeta \rceil$ points; therefore, for sufficiently small ζ, it can definitely be done with cardinality bounded by $2\pi/\zeta$. To proceed from S^k to S^{k+1}, observe that we can inductively cover the hypercylinder $S^k \times [-\pi/2, \pi/2]$ by $(c/\zeta)^k \times \lceil \pi/(2\zeta) \rceil$ points, so we can definitely cover that hypercylinder by $(c/\zeta)^{k+1}$ points. Finally, we obtain a covering of S^{n+1} from our covering of the hypercylinder via the contractive map $(\xi; t) \mapsto (\xi \cos t; \sin t)$, where ξ is an n-vector and t is a scalar. □

Let us apply Lemma 1; fix an ζ-net $\{v_k\}_1^M$ of $\mathbb{S}^{2D-1}$, with cardinality $M \leq (c/\zeta)^{2D-1}$. Equation (5) becomes:

$$\lambda(\Psi) = \max_{1 \leq k \leq M} \left\{ \sup_{v \in B'(v_k, \zeta)} |\langle \Psi v, v \rangle| \right\}, \tag{6}$$

where $B'(v_k, \zeta)$ denotes the intersection, in $\mathbb{C}^D$, of $\mathbb{S}^{2D-1}$ with the ball with center v_k and radius ζ.

Lemma 2 *If $v, v_k \in \mathbb{S}^{2D-1}$, and $\|v - v_k\| \leq \zeta$, then $| \langle \Psi v, v \rangle - \langle \Psi v_k, v_k \rangle | \leq 2\zeta$.*

Proof: Let $\nu = v - v_k$. Since Ψ is an average of unitary matrices, its matrix norm is bounded by 1; therefore:

$$\begin{aligned} | \langle \Psi(v_k + \nu), v_k + \nu \rangle - \langle \Psi v_k, v_k \rangle | &= | \langle \Psi(v_k + \nu), \nu \rangle + \langle \Psi \nu, v_k \rangle | \\ &\leq |\langle \Psi(v_k + \nu), \nu \rangle| + |\langle \Psi \nu, v_k \rangle| \\ &\leq \|\Psi\| \cdot \|v_k + \nu\| \cdot \|\nu\| + \|\Psi\| \cdot \|\nu\| \cdot \|v_k\| \\ &\leq 1 \cdot 1 \cdot \zeta + 1 \cdot \zeta \cdot 1 = 2\zeta \end{aligned}$$

□

Applying Lemma 2, we pass from equation (6) to:

$$\lambda(\Psi) \leq \max_{1 \leq k \leq M} |\langle \Psi v_k, v_k \rangle| + 2\zeta. \tag{7}$$

Next, we isolate the v_k as follows:

$$\Pr[\lambda(\Psi) \geq \delta] \leq \Pr[\exists k \text{ s.t. } |\langle \Psi v_k, v_k \rangle| + 2\zeta \geq \delta] \leq \sum_{k=1}^{M} \Pr[|\langle \Psi v_k, v_k \rangle| \geq \delta - 2\zeta] \tag{8}$$

Our next objective is to study the distribution of $\langle \Psi v, v \rangle$ for fixed v. We shall show that we can uniformly bound this distribution; furthermore, our result will be independent of v, so we will be able to apply it to all terms of the above sum.

2.4 Analysis along a single vector

Let v be a fixed unit vector in $\mathbb{C}^D$. We can split our inner product into:

$$\langle \Psi v, v \rangle = \left\langle \frac{1}{|S|} \left(\sum_{s \in S} \rho(s) \right) v, v \right\rangle = \frac{1}{|S|} \sum_{s \in S} \langle \rho(s) v, v \rangle. \tag{9}$$

Regroup the terms in the sum as follows: recall that S was formed by accumulating blocks from our initial partition $\mathcal{P}(G)$; group the $\rho(s)$ according to those blocks. We will obtain the form:

$$\sum_{s\in S}\rho(s)=\sum_{i=1}^{N}H_i,$$

where the (distinct) H_i are Hermitian matrices that are either of the form $\rho(s)$ or $(\rho(t)+\rho(t^{-1}))$, where the s and t have the same meanings as in Section 2.1. Hence equation (9) becomes:

$$\langle\Psi v,v\rangle=\frac{1}{|S|}\sum_{i=1}^{N}\langle H_iv,v\rangle\,. \tag{10}$$

To simplify notation, let us define a functional f_v on the set of blocks in $\mathcal{P}(G)$; for a block B, let

$$f_v(B)=\left\langle\left(\sum_{s\in B}\rho(s)\right)v,v\right\rangle$$

Observe that we defined our blocks so that each $\sum_{s\in B}\rho(s)$ would be Hermitian, because ρ is a unitary representation. Therefore, f_v is a real functional, parameterized by v. Furthermore, since we formed S by sampling blocks without replacement, our (stochastic) sum in equation (10) is actually just:

$$\langle\Psi v,v\rangle=\frac{N}{|S|}\overline{I}_N, \tag{11}$$

where $\overline{I}_N$ is the sample mean of N elements sampled without replacement from the population $\mathcal{I}=\{f_v(B)\}_{B\in\mathcal{P}(G)}$, with respect to the uniform distribution. This population has a special property:

Lemma 3 *The mean of the population $\mathcal{I}=\{f_v(B)\}_{B\in\mathcal{P}(G)}$ is zero.*

Proof: This calculation can be performed explicitly. Let $\phi=\sum_{s\in G}\rho(s)$. Then:

$$\sum_{B\in\mathcal{P}(G)}f_v(B)=\sum_{s\in G}\langle\rho(s)v,v\rangle=\left\langle\left(\sum_{s\in G}\rho(s)\right)v,v\right\rangle=\langle\phi v,v\rangle \tag{12}$$

Now let $\rho^*:G\mapsto(\mathrm{id})_D$ be the trivial representation on $\mathrm{Mat}_{D\times D}(\mathbb{C})$. For every $s\in G$, $\rho^*(s)\circ\phi=\phi\circ\rho(s)$. By Schur's Lemma [7], since ρ was assumed nontrivial, we must have $\phi\equiv 0$. Therefore, the mean is zero. □

2.5 Specialization to Abelian case

We now proceed to prove Theorem 2. In this section, we shall finally make the assumption that G is Abelian; however, the only things we derive from the Abelian property are that all irreducible representations of G are one-dimensional, and that there are $|G|$ of them (i.e. $R=|G|$). To obtain a large-deviation bound on $\langle\Psi v,v\rangle$, we use one of Hoeffding's inequalities from [4]:

Fact 2 *(Hoeffding, 1963) Let a population $\mathcal{C}$ consist of n real values $c_1,c_2,\ldots,c_n$, such that all of the c_i lie in the interval $[-2,2]$ and their mean is zero. Let $\overline{X}$ denote the sample mean of N elements taken without replacement from $\mathcal{C}$. Then:*

$$\Pr\left[\overline{X}\geq t\right]\leq e^{-\frac{1}{2}Nt^2} \tag{13}$$

We shall use the following immediate corollary of that fact:

Corollary 1 *Start with the same setup as Fact 2. Now suppose that k is a positive constant bounded by 1. Then:*

$$\Pr\left[k|\overline{X}|\geq t\right]\leq 2e^{-\frac{1}{2}Nt^2} \tag{14}$$

Since our representation is unitary and $|B| \leq 2$, we immediately satisfy the condition $f_v(B) \in [-2, 2]$. Also, by definition of S, $N \leq |S|$, so our k in Corollary 1 is bounded by 1. Lemma 3 secures the remaining condition of Fact 2; therefore, we have:

$$\begin{aligned}
\Pr\left[|\langle \Psi v, v\rangle| \geq \delta - 2\zeta\right] &= \Pr\left[\frac{N}{|S|}\left|\overline{I}_N\right| \geq \delta - 2\zeta\right] \leq 2e^{-\frac{1}{2}N(\delta-2\zeta)^2} \\
\Pr\left[\lambda(\Psi_i) \geq \delta\right] &\leq \sum_{k=1}^{M} \Pr\left[|\langle \Psi v_k, v_k\rangle| \geq \delta - 2\zeta\right] \leq M \cdot 2e^{-\frac{1}{2}N(\delta-2\zeta)^2} \\
\Pr\left[\mu(X(G,S)) \geq \delta\right] &= \Pr\left[\max_{i\neq 1}\{\lambda(\Psi_i)\} \geq \delta\right] \leq \sum_{i=2}^{R} \Pr\left[\lambda(\Psi_i) \geq \delta\right] \leq (R-1)M \cdot 2e^{-\frac{1}{2}N(\delta-2\zeta)^2}
\end{aligned}$$

Since our dimension D is 1, Lemma 1 tells us that $M \leq c/\zeta$. Using the fact that $R = |G|$,

$$\begin{aligned}
\Pr\left[\mu(X(G,S)) \geq \delta\right] &\leq (|G|-1)\left(\frac{c}{\zeta}\right) \cdot 2e^{-\frac{1}{2}N(\delta-2\zeta)^2} \\
&\leq |G|\left(\frac{4c}{\delta}\right) \cdot 2e^{-\frac{1}{8}N\delta^2} \quad \text{by choosing the particular value of } \delta/4 \text{ for } \zeta
\end{aligned}$$

Therefore, for any finite Abelian[1] group G, we can accomplish $\Pr\left[\mu(X(G,S)) \geq \delta\right] \leq \epsilon$ for all $N \geq C_0(\epsilon, \delta)\log|G|$, where the multiplicative constant C_0 depends only on ϵ and δ.

□

2.6 Non-Abelian Case

In this section, we use our results from Section 2.4 to motivate Conjecture 1. We drop Section 2.5's assumption that $D = 1$, and continue from where we left off at the end of Section 2.4. We first show that Conjecture 1 is a consequence of the following conjecture:

Conjecture 2 *Let $\overline{I}_N$ have the same meaning as in equation (11). Then there exist some positive constants α and β, independent of N, D, and G, such that:*

$$\Pr\left[\left|\overline{I}_N\right| \geq t\right] \leq \alpha e^{-\beta DNt^2} \tag{15}$$

From a heuristic point of view, this means that D and N should appear as a fundamental unit; that is, our averages really have sample size DN, not just N.

To establish the connection between our two conjectures, we continue with the analogue of what we did with Corollary 1.

$$\begin{aligned}
\Pr\left[|\langle \Psi v, v\rangle| \geq \delta - 2\zeta\right] &= \Pr\left[\frac{N}{|S|}\left|\overline{I}_N\right| \geq \delta - 2\zeta\right] \leq \alpha e^{-\beta DN(\delta-2\zeta)^2} \\
\Pr\left[\lambda(\Psi_i) \geq \delta\right] &\leq \sum_{k=1}^{M} \Pr\left[|\langle \Psi v_k, v_k\rangle| \geq \delta - 2\zeta\right] \leq M \cdot \alpha e^{-\beta DN(\delta-2\zeta)^2} \\
&\leq \left(\frac{c}{\zeta}\right)^{(2D-1)} \cdot \alpha e^{-\beta DN(\delta-2\zeta)^2} \quad \text{(by Lemma 1)} \\
&\leq \alpha \exp\left\{(2D-1)\log\left(\frac{c}{\zeta}\right) - \beta DN(\delta-2\zeta)^2\right\}
\end{aligned}$$

[1] In fact, more is true; our analysis remains valid as long as D is uniformly bounded by some constant, over all possible irreducible representations of the groups under consideration. For example, our proof will still work if we include the family of Dihedral groups, because all of their D are bounded by 2.

Now, as long as $\zeta \leq c$ and $\beta N(\delta - 2\zeta)^2 \geq 2\log(c/\zeta)$, we have:

$$\begin{aligned}
\Pr\left[\lambda(\Psi_i) \geq \delta\right] &\leq \alpha \exp\left\{D\left[2\log\left(\frac{c}{\zeta}\right) - \beta N(\delta - 2\zeta)^2\right]\right\} \\
&\leq \alpha \exp\left\{2\log\left(\frac{c}{\zeta}\right) - \beta N(\delta - 2\zeta)^2\right\} \\
&\leq \alpha \left(\frac{c}{\zeta}\right)^2 e^{-\beta N(\delta - 2\zeta)^2} \\
\Pr\left[\mu(X(G,S)) \geq \delta\right] \leq \sum_{i=2}^{R} \Pr\left[\lambda(\Psi_i) \geq \delta\right] &\leq (R-1)\alpha \left(\frac{c}{\zeta}\right)^2 e^{-\beta N(\delta - 2\zeta)^2}
\end{aligned}$$

Once again, the arbitrary choice of $\zeta = \delta/4$ yields the fact that for any finite group G, we have $\Pr\left[\mu(X(G,S)) \geq \delta\right] \leq \epsilon$ for all $N \geq C(\epsilon, \delta)\log R$, where C depends[2] only on ϵ and δ.

□

3 Support for Conjecture 2

We conclude this paper by providing two heuristic arguments that support our conjecture. The first uses the Central Limit Theorem, and the second studies the nature of the population $\mathcal{I} = \{f_v(B)\}_{B \in \mathcal{P}(G)}$.

3.1 Central Limit Theorem

We already know from Lemma 3 that the mean of $\mathcal{I}$ is zero. Let us estimate the variance. Suppose that the partition $\mathcal{P}(G)$ is as in equation (1). We then have:

$$\begin{aligned}
\sigma^2 &= \frac{1}{|\mathcal{I}|} \sum_{B \in \mathcal{P}(G)} f_v(B)^2 \\
&= \frac{1}{a+b}\left(\sum_{i=1}^{a} \langle \rho(s_i)v, v\rangle^2 + \sum_{i=1}^{b} \left\langle (\rho(t_i) + \rho(t_i^{-1}))v, v\right\rangle^2\right) \\
&= \frac{1}{a+b}\left(\sum_{i=1}^{a} \langle \rho(s_i)v, v\rangle^2 + \sum_{i=1}^{b} \langle \rho(t_i)v, v\rangle^2 + \sum_{i=1}^{b} \left\langle \rho(t_i^{-1})v, v\right\rangle^2 + 2\sum_{i=1}^{b} \langle \rho(t_i)v, v\rangle \left\langle \rho(t_i^{-1})v, v\right\rangle\right) \\
&= \frac{1}{a+b}\left(\sum_{g \in G} \langle \rho(g)v, v\rangle^2 + 2\sum_{i=1}^{b} \langle \rho(t_i)v, v\rangle \left\langle \rho(t_i^{-1})v, v\right\rangle\right)
\end{aligned}$$

Let us perform a unitary change-of-basis so that v becomes the first basis element. Let the new matrices be $\rho'(g) = (r_{ij}(g))$; we shall use the following corollary to Schur's Lemma from [7]:

Fact 3 *(from Serre)*

$$\sum_{g \in G} r_{11}(g) r_{11}(g^{-1}) = \frac{|G|}{D}$$

Our change of basis transforms our last equation into:

[2]The constraint $\beta N(\delta - 2\zeta)^2 \geq 2\log(c/\zeta)$ does not cause any problems because it is independent of G.

$$\begin{aligned}
\sigma^2 &= \frac{1}{a+b}\left(\sum_{g\in G} r_{11}(g)^2 + 2\sum_{i=1}^{b} r_{11}(t_i)r_{11}(t_i^{-1})\right) \\
&= \frac{1}{a+b}\left(\sum_{g\in G} r_{11}(g)^2 + \sum_{g\in G}|r_{11}(g)|^2 - \sum_{i=1}^{a}|r_{11}(s_i)|^2\right) \quad \text{since } r_{11}(g^{-1}) = \overline{r_{11}(g)} \\
&\leq \frac{1}{a+b}\left(\sum_{g\in G}|r_{11}(g)|^2 + \sum_{g\in G}|r_{11}(g)|^2 - \sum_{i=1}^{a}|r_{11}(s_i)|^2\right) \\
&\leq \frac{2|G|}{D(a+b)}
\end{aligned}$$

Applying the Central Limit Theorem (with a correction factor because we are sampling without replacement), we obtain that:

$$\overline{I}_N \approx \text{Normal}\left(0, \frac{2|G|}{DN(a+b)}\frac{a+b-N}{a+b-1}\right) \tag{16}$$

Hence the distribution of $\overline{I}_N$ has (approximately) thinner tails than Normal$(0, 4/(ND))$. Use the following estimate from [3]:

Fact 4 *(from Feller) Let Z be the standard Normal random variable. For every $x \geq 0$:*

$$\Pr\left[|Z| \geq x\right] \leq \min\left\{1, \frac{1}{x}e^{-x^2/2}\right\}$$

Therefore, there exists a positive constant α such that $\Pr\left[|Z| \geq x\right] \leq \alpha e^{-x^2/2}$. This implies that there exists a positive constant β for which:

$$\Pr\left[\left|\overline{I}_N\right| \geq t\right] \leq \alpha e^{-\beta DNt^2},$$

as long as our Central Limit approximation is sufficiently tight.

3.2 Population $\mathcal{I}$

In the Abelian case, we used Hoeffding's Inequality to obtain our large-deviation bound. That inequality only needed the boundedness of the population $\mathcal{I}$; it did not take into account the fact that the variance scaled as $1/D$. This scaling factor is quite interesting, because it hints that the individual elements of $\mathcal{I}$ may themselves be D-wise means, from some deeper population. Then $\overline{I}_N$ would be a DN-wise mean, which would support our hypothesis that DN should appear as a fundamental unit in a large-deviation bound.

Let us express $\mathcal{I}$ in an even more suggestive form. Consider some $f_v(B) = \left\langle\left(\sum_{s\in B}\rho(s)\right)v, v\right\rangle$, and let $H_B = \sum_{s\in B}\rho(s)$. Note that by construction, H_B must be Hermitian, hence diagonalizable with an orthonormal basis of eigenvectors, $\{e_k\}_1^D$. Let the corresponding eigenvalues be $\{\lambda_k\}_1^D$. Express v in that basis; suppose that $v = \sum c_k e_k$. Then $f_v(B) = \sum c_k^2\lambda_k$, and furthermore, since the basis was orthonormal, $\sum c_k^2 = 1$. Therefore, $f_v(B)$ is a D-wise weighted mean of the λ_k, which also happen to be bounded by $[-2, 2]$ because $|B| \leq 2$ and $\rho(s)$ are unitary. The weights $\{c_k\}_1^D$ are not necessarily balanced, however, so it remains to study how well-balanced they are under large D.

4 Acknowledgements

Many thanks to the Marshall family for their generous financial support, and to the Student Undergraduate Research Fellowship office at Caltech for arranging this research project.

References

[1] N. Alon and Y. Roichman, *Random Cayley Graphs and Expanders*, Random Structures and Algorithms, 5 (1994), pp. 271–284.

[2] N. Alon, J. Spencer, and P. Erdos, *The Probabilistic Method*, Wiley, 1992, pp. 119–125.

[3] W. Feller, *An Introduction to Probability Theory and Its Aplications*, Wiley, 1968, vol. 1, p. 175.

[4] W. Hoeffding, *Probability Inequalities For Sums Of Bounded Random Variables*, Journal of the American Statistical Association, 58 (1963), pp. 13–30.

[5] A. Lubotzky, *Cayley Graphs: Eigenvalues, Expanders and Random Walks*, in Surveys in Combinatorics 1995, P. Rowlinson ed., Cambridge University Press, 1995, pp. 155–189.

[6] A. Lubotzky, R. Phillips, and P. Sarnak, *Ramanujan Graphs*, Combinatorics, 8 (1988), pp. 261–277.

[7] J. Serre, *Linear Representations of Finite Groups*, Springer, 1977.

[8] B. Simon, *Representations of Finite and Compact Groups*, American Mathematics Society, 1996.