Guest Column: Proving expansion in three steps[1]

Amir Yehudayoff [2]

Abstract

This text is meant to be an introduction to a recent strategy introduced by Bourgain and Gamburd (following a work of Helfgott) for proving graph-expansion. The strategy is designed for graphs H that are defined using some underlying group G.

The strategy consists of three steps, which, in Sarnak's terminology, correspond to the three steps of a chess game: opening, middle-game and endgame. In the opening, the objective is to prove that the girth of H is logarithmic. In the middle-game, the goal is to prove a product-growth theorem for subsets of G. The endgame consists of establishing a "mixing property" for G. There are two methods for proving a mixing property: using pairwise independence and using basic representation theory.

1 Introduction

Expanders are graphs with good connectivity properties. They turn out to be extremely useful in many areas of research. This column does not discuss the vast applications of expanders, but rather focuses on describing an approach for constructing such graphs, or, more precisely, proving that a given graph is an expander. For a detailed survey of applications and properties of expanders, see [11].

There are, in general, three strategies for building expanders or proving that a given graph is an expander. The historically first strategy was suggested by Margulis [15] and is algebraic in nature[3] (see also [14]). The second approach is combinatorial and iterative in nature. It was pioneered by the zig-zag product of Reingold, Vadhan and Wigderson [17]. The third approach is analytic and uses additive combinatorics. This approach was introduced in the works of Bourgain and Gamburd [2, 3] that use ideas from the works of Helfgott [10] and Sarnak and Xue [19]. This approach is quite general and enables to prove expansion in cases that were not known before.

The aim of this text is to be a tutorial to the third. We choose simplicity over abstractness: Statements far more general than the ones we provide here are available in the literature, e.g. [16, 5]. Although this text is self-contained, it is only meant to serve as an introduction to this line of research, and some parts are not fully explained.

[2] Department of Mathematics, Technion-IIT, Haifa, Israel. amir.yehudayoff@gmail.com. Horev fellow – supported by the Taub foundation. Supported by grants from ISF and BSF.

[3] Gabber and Galil [8] in fact found a way to use elementary harmonic analysis to prove that one of Margulis' constructions works. Davidoff, Sarnak and Valette [6] gave a proof that a different graph is an expander using elementary counting and basic representation theory.

We start with a formal introduction to expander graphs. We later discuss two generic ways in which groups define graphs. Most of the text provides a glance into the three-step proof of Bourgain and Gamburd [3, 2, 10, 4].

1.1 Expander graphs

We provide two formal definitions of expander graphs and discuss a third one. Each definition provides a different point of view of expanders. There are more equivalent definitions (see [11]) that are extremely important in applications but we shall not consider here. The term "an expander graph" always refers to an infinite family of graphs $\{H_n\}$ of increasing sizes. As we shall see, the definition is not interesting for a single graph.

The first property required from a family of expander graphs is to have constant degree. For simplicity and concreteness, all families of expander graphs we discuss are d-regular, for a constant d independent of n.

We define expanders via "edge expansion" and state that it is equivalent to a spectral definition. Let $H = (V, E)$ be a d-regular graph. The *edge expansion* $\mathsf{h}(H)$ of H is the smallest normalized size of a cut in H: For every $A \subset V$, denote by $E(A, \overline{A})$ the set of edges with one endpoint in A and one endpoint not in A. Define

$$\mathsf{h}(H) = \min \left\{ \frac{|E(A, \overline{A})|}{d|A|} : A \subset V,\ 0 < |A| \leq \frac{|V|}{2} \right\}.$$

Definition (expander)**.** *The graph H is called an* expander*, if there exists $\varepsilon > 0$, a constant independent of the size of H, so that* $\mathsf{h}(H) \geq \varepsilon$*.*

The *spectral gap* $\gamma(H)$ of H is the spectral gap of the normalized adjacency matrix of H: The graph H defines a normalized adjacency matrix M by

$$M_{v,u} = \begin{cases} 1/d & \{v, u\} \in E, \\ 0 & \{v, u\} \notin E. \end{cases}$$

The matrix M is symmetric so it has $n = |V|$ real eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$. Since it is normalized, all eigenvalues are in $[-1, 1]$. Always $\lambda_1 = 1$, corresponding to the all-ones eigenvector $\overline{1}$. When H is connected, this is the only eigenvector of eigenvalue 1. When H is not bi-partite[4], $\lambda_n > -1$. Define the *spectral gap* of H as

$$\gamma(H) = 1 - \lambda_2.$$

The following well-known theorem shows that expansion can be cast using $\gamma(H)$ too (see [11] and references within).

Theorem 1 (Cheeger's inequality)**.** *For every d-regular graph H,*

$$\frac{\gamma(H)}{2} \leq \mathsf{h}(H) \leq \sqrt{2\gamma(H)}.$$

[4] When H is bi-partite, a slightly different discussion is required.

This fundamental inequality translates algebraic information to a combinatorial one and vice versa. Specifically, it shows that we could have defined "expander" by $\gamma(H) \geq \varepsilon$, for some constant $\varepsilon > 0$. We shall not provide a proof here, but we shortly discuss the proof idea, as it may help to gain intuition. There are two inequalities to prove.

To prove the left inequality assume that $\gamma(H)$ is greater than zero. Consider a set A of vertices with minimal-size normalized cut. Define the vector x in $\mathbb{R}^V$ as $x_v = 1 - |A|/n$ when $v \in A$ and $x_v = -|A|/n$ otherwise. The idea is to consider $\langle x, Mx \rangle$ in two ways. On one hand, $\langle x, Mx \rangle$ measures the cut-size of A. On the other hand, when decomposing x as a linear combination of the orthonormal eigenvectors of M, since x is orthogonal to $\overline{1}$, the inequality follows.

The right inequality is harder to prove. It states that if $\gamma(H)$ is close to zero, then the edge expansion is small, namely, that there is a set A of vertices with few edges in $E(A, \overline{A})$. The idea is to consider x, the eigenvector of M that corresponds to $\lambda_2 = 1 - \gamma(H)$, and use it to define a set A. The way to define A using x is to use some threshold $t \in \mathbb{R}$ and choose A as the set of vertices v so that $x_v \geq t$.

A third definition of expanders is based on entropy of random walks. A *(simple) random walk* on a graph G is a random sequence of vertices $v_0, v_1, v_2, \ldots$ so that v_{t+1} is a uniform random neighbor of v_t, independent of all previous choices. The distribution of the starting vertex v_0 is called the *initial* distribution. The distribution of v_t is denoted in this text as μ_t. A *stationary* distribution for a regular graph is uniform: if the initial distribution is the uniform distribution u, then μ_1 is also uniform, μ_2 is too and so forth.

The second law of thermodynamics says that entropy increases with time. The entropy in our context is measured as $1/\left\|\mu_t\right\|_2^2$ (this is the exponential of the Rényi entropy). This entropy indeed does not decrease with time (other "types" of entropy may decrease with time). The maximal entropy is achieved by the uniform distribution $1/\left\|\mathsf{u}\right\|_2^2 = |G|$. A third possible definition of an "expander graph" can be as a graph on which the entropy increments are bounded from below so that in logarithmic time the entropy becomes very close to maximal (regardless of the initial distribution). We shall use this intuition as well.

Before we provide explicit examples of expander graphs, we explain how can one use groups to define graphs. The expanders we discuss are group-based graphs.

1.2 Group-based graphs

Consider a group G. A key example we consider is the group of two by two matrices with determinant one over some prime field $\mathbb{F}$,

$$\mathsf{SL}_2(\mathbb{F}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a, b, c, d \in \mathbb{F}, \ ad - bc = 1 \right\}.$$

The group operation in this example is simply product of matrices.

We start by describing the family of *Cayley* graphs G defines (see also [11]): Let S be a finite subset of G. Denote $S^{-1} = \{g^{-1} : g \in S\}$. The vertices of the Cayley graph $\mathsf{Cay}(G, S)$ are the elements of G and its edges are of the form (g, sg) for all g in G and s in $S \cup S^{-1}$ (it is hence an undirected graph). Cayley graphs are highly symmetric graphs, specifically, they are vertex transitive. If the set S generates G, $\langle S \rangle = G$, then $\mathsf{Cay}(G, S)$ is connected (and vice versa).

As an example consider the subset of $\mathsf{SL}_2(\mathbb{F})$, for $\mathbb{F}$ large but finite,

$$S = \left\{ \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \right\}. \tag{1}$$

The graph $\mathsf{Cay}(\mathsf{SL}_2(\mathbb{F}), S)$ is connected, of size $|\mathsf{SL}_2(\mathbb{F})| \sim |\mathbb{F}|^3$ and is four-regular.

A more general family of graphs is that of *Schreier diagrams*: The group G can *act* on some set X. An *action* of G on X is a homomorphism from g to functions from X to X, namely, every g in G defined a map[5] $g(\cdot)$ from X to X so that for every g' in G,

$$g(g'(\cdot)) = (gg')(\cdot).$$

Specifically, $g(\cdot)$ is invertible with inverse $(g^{-1})(\cdot)$. Given a finite subset S of G, we can again use it to define a graph, $\mathsf{Sch}(G, S, X)$. The vertex set of the graph is X and the edges are of the form $(x, s(x))$ for x in X and s in $S \cup S^{-1}$. A Cayley graph is a Schreier diagram w.r.t. the action of G on itself by left-multiplication: $g(h) = gh$.

Another example of a group action is the Möbius action of $\mathsf{SL}_2(\mathbb{F})$ on the projective line $P(\mathbb{F}) = \mathbb{F} \cup \{\infty\}$ defined by

$$g(x) = \frac{ax+b}{cx+d}, \text{ where } g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Here one can interpret operations using infinity in the natural way (e.g. $1/\infty = 0$). It is easy to verify that this is indeed an action:

$$g(g'(x)) = \frac{a\frac{a'x+b'}{c'x+d'} + b}{c\frac{a'x+b'}{c'x+d'} + d} = \frac{x(aa' + bc') + ab' + bd'}{x(ca' + dc') + cb' + dd'} = (gg')(x).$$

When S is the set defined in (1), the diagram $\mathsf{Sch}(\mathsf{SL}_2(\mathbb{F}), S, P(\mathbb{F}))$ is connected, has $|\mathbb{F}|+1$ vertices and is four-regular.

There are two useful families of actions that we consider: transitive and two-transitive. An action of G on X is *transitive* if for every x_1, x_2 in X, there is g in G so that $gx_1 = x_2$. A simple example is the action of G on itself by left-multiplication. A Schreier diagram with S that generates G is connected iff the action is transitive. An action is *two-transitive* if the action it defines on distinct pairs in X is transitive, that is, for every $x_1 \neq x_2$ and $x_3 \neq x_4$ in X, there is g in G so that $g(x_1) = x_3$ and $g(x_2) = x_4$. The family of two-transitive actions is much smaller than that of transitive actions. Two well-known examples of two-transitive actions is (i) the affine group acting on the underlying field and (ii) the Möbius action defined above. Two-transitive actions define pairwise independent hash functions, and the use of two-transitivity here is similar to the way pairwise independent hashing is typically used.

We explain why the Möbius action is two-transitive (it is in fact three-transitive). Let $x_1 \neq x_2$ be two distinct elements of $\mathbb{F}$. We find a matrix $g \in \mathsf{SL}_2(\mathbb{F})$ that maps the pair (x_1, x_2) to the pair $(0, 1)$: First choose $g_1 \in \mathsf{SL}_2(\mathbb{F})$ that maps (x_1, x_2) to $(0, x_2 - x_1)$,

$$g_1 = \begin{pmatrix} 1 & -x_1 \\ 0 & 1 \end{pmatrix}.$$

[5] This is a slight abuse of notation.

Second choose $g_2 \in \mathsf{SL}_2(\mathbb{F})$ that maps $(0, x_2 - x_1)$ to $(0, 1)$,

$$g_2 = \begin{pmatrix} 1 & 0 \\ 1 - 1/(x_2 - x_1) & 1 \end{pmatrix}.$$

So the matrix $g = g_2 g_1$ maps (x_1, x_2) to $(0, 1)$ as wished (a similar construction works when one of x_1, x_2 is infinity). Since every group element has an inverse, this implies that the action is indeed two-transitive: if g maps (x_1, x_2) to $(0, 1)$, and g' maps (x_3, x_4) to $(0, 1)$, then $g'^{-1}g$ maps (x_1, x_2) to (x_3, x_4).

1.3 Main example and some comments

The historically first construction of an explicit expander graph is a Schreier diagram that is due to Margulis [15]. We shall not prove that Margulis' construction works, but we shall explain how to prove that a similar construction works. Our main example is a Schreier diagram with the group $G = \mathsf{SL}_2(\mathbb{F})$ where $\mathbb{F}$ is a prime field of size p, the set S defined in (1), and the Möbius action of G on $X = P(\mathbb{F})$. A similar proof shows that the Cayley graph defined by these G, S is an expander. We choose the Schreier diagram example to emphasize the generality of the approach, and since some parts of the proof are only relevant to Schreier diagrams (e.g. there is no two-transitive action of G on itself). The proof generalizes to other cases as well.

We give a simple example that emphasizes the strength of this proof technique compared to previous algebraic techniques. Consider three sets of generators S_1, S_2, S_3 defined by

$$S_i = \left\{ \begin{pmatrix} 1 & i \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ i & 1 \end{pmatrix} \right\},\ i \in \{1, 2, 3\}.$$

So S_2 is S we chose above. Lubotzky 1-2-3 question [13] asks whether all three $\mathsf{Cay}(G, S_i)$ are expanders for $G = \mathsf{SL}_2(\mathbb{F})$ and a large prime field $\mathbb{F}$. The "algebraic" method for proving expansion (specifically Selberg's theorem) implies that $\mathsf{Cay}(G, S_1), \mathsf{Cay}(G, S_2)$ are expanders, but does not imply that $\mathsf{Cay}(G, S_3)$ is an expander. In fact, that $\mathsf{Cay}(G, S_2)$ is an expander can be proved in an elementary manner by counting cycles in the graph and using dimension-multiplicity (see the endgame, Section 5.2) as was done in [19, 6]. In [2] Bourgain and Gamburd, when introducing the technique we survey here, solved the question and proved that the answer is yes, all three graphs are expanders.

Before moving to proofs, we describe two applications of this machinery. The first is related to quantum computing. Bourgain and Gumburd [2] used this method to prove expansion in the group of two by two unitary matrices SU(2). This was used by Bourgain to show optimal convergence of the Solovay-Kitaev algorithm (see [7] and references within). The second is a construction of monotone expanders. In [4] expansion in $\mathrm{SL}_2(\mathbb{R})$, the group of two by two real matrices with determinant one, was proved, and then used to give the only construction known of monotone expanders.

2 The game

We describe an argument, following Bourgain and Gamburd [2, 3], for proving that $\mathsf{Sch}(G, S, X)$, $\mathsf{Cay}(G, S)$ are expanders. The proof consists of three steps. In Sarnak's terms [18], these three steps correspond to the three steps of a chess game: opening, middle-game and endgame. As in chess, each step in the proof has a different objective and hence strategy.

To prove expansion using this approach, we need to prove three lemmas (or variants of them) that correspond to the three steps above. In this section, we state the lemmas for general G, S, X with the relevant conditions, and show how to combine the lemmas to a proof. In later sections, we discuss each lemma separately, with a focus on our main example ($G = \mathsf{SL}_2(\mathbb{F})$, S as in (1), and the Möbius action on $X = P(\mathbb{F})$).

Opening. In the opening stage, we just argue that $\mathsf{Cay}(G, S)$ has large girth (the girth of a graph is the length of the shortest cycle in it). As we shall see, except from the trivial requirement $\langle S \rangle = G$, this is the only step that uses properties of S.

Lemma 2. *There exists a universal constant $c_1 > 0$ so that the graph $\mathsf{Cay}(G, S)$ has girth at least*[6] $c_1 \log |G|$.

Middle-game. In the middle-game, a product growth statement is required (the first instantiation of such a lemma was proved by Helfgott [10]). This step just depends on the group (it does not use S nor the action).

Lemma 3. *There exists a constant $c > 0$ so that the following holds. For every $\delta > 0$, there exists $\varepsilon > 0$ so that for every $A \subset G$ of size $|A| \leq |G|^{1-\delta}$ so that $\langle A \rangle = G$,*

$$|A \cdot A \cdot A| \geq c|A|^{1+\varepsilon},$$

where $A \cdot A \cdot A = \{a_1 a_2 a_3 : a_1, a_2, a_3 \in A\}$.

Endgame. The last step requires proving a mixing property. This is the only step that is related to the action of G on X.

Lemma 4. *There exists a universal constant $c_2 > 0$ so that for every probability distribution μ on G and for every $f : X \to \mathbb{R}$ so that $\sum_{x \in X} f(x) = 0$,*

$$\|\mu * f\|_2^2 \leq \frac{|G|}{|G|^{c_2}} \|\mu\|_2^2 \, \|f\|_2^2 \, ,$$

where $*$ is convolution: $\mu * f : X \to \mathbb{R}$ defined by

$$(\mu * f)(x) = \sum_{g \in G} \mu(g) f(g^{-1}(x)).$$

These are the three lemmas that guarantee that $H = \mathsf{Sch}(G, S, X)$ is an expander. The goal will (thus) be to prove that $\lambda = \lambda(H) = 1 - \gamma(H)$ is uniformly bounded away from one. We shall also use the intuition that we wish to prove that the entropy of a random walk increases rapidly and becomes close to maximal in logarithmic time. Although we are eventually interested in $\mathsf{Sch}(G, S, X)$, we shall first consider the behavior of a random walk on $\mathsf{Cay}(G, S)$.

[6] Logarithms in this text are base two, unless otherwise stated.

Definition (random walk on H)**.** *Define μ_t to be the distribution of a simple random walk on $\mathsf{Cay}(G,S)$ of length t started at the identity: μ_0 is supported on the identity of G,*

$$\mu_1(g) = \begin{cases} \frac{1}{|S|} & g \in S \cup S^{-1} \\ 0 & g \notin S \cup S^{-1} \end{cases}$$

and for $t > 1$,

$$\mu_t = \mu_1 * \mu_{t-1}.$$

Let $f : X \to \mathbb{R}$ be the second eigenvector of the normalized adjacency matrix of H, that is, $\mu_1 * f = \lambda f$ and iterating for all integers $t > 0$,

$$\mu_t * f = \lambda^t f.$$

We shall argue that for $t_3 \le C \log |G|$, with $C > 0$ a universal constant,

$$\lambda^{2t_3} \|f\|_2^2 = \|\mu_{t_3} * f\|_2^2 \le \frac{1}{|G|^{c_2/c}} \|f\|_2^2, \tag{2}$$

with $c_2 > 0$ the constant from the endgame lemma. The left equality follows by choice of f and the right inequality will follow using the three lemmas. This clearly completes the proof:

$$\lambda \le 2^{-c_2/(2C)} < 1.$$

Combining the three steps. It remains to explain how (2) follows from the three lemmas. The proof, naturally, consists of three parts as below.

Before describing the three parts, we explain, in high level, how the proof that $\mathsf{Cay}(G,S)$ is an expander goes (for $\mathsf{Sch}(G,S,X)$ details follow). Consider the entropy $1/\|\mu_t\|_2^2$. The reader may think of μ_t as being uniform over a set of size $1/\|\mu_t\|_2^2$. Our goal is to show that in logarithmic time the entropy is nearly maximal. The proof considers four times: $0 = t_0 \le t_1 \le t_2 \le t_3$ where $t_i = \alpha_i \log n$ with some constants $0 = \alpha_0 \le \alpha_1 \le \alpha_2 \le \alpha_3$. In time t_0 the entropy is minimal $1/\|\mu_{t_0}\|_2^2 = 1$. The opening will imply that the entropy at time t_1 is order $|G|^{\varepsilon}$ for some constant $\varepsilon > 0$. The middle-game will imply that at time t_2 the entropy is actually $|G|^{1-\varepsilon}$. The endgame, finally, will imply that at time t_3 the entropy is order $|G|$. Schematically,

$$1/\|\mu_{t_0}\|_2^2 = 1 \xrightarrow{\text{opening}} 1/\|\mu_{t_1}\|_2^2 = |G|^{\varepsilon} \xrightarrow{\text{middle-game}} 1/\|\mu_{t_2}\|_2^2 = |G|^{1-\varepsilon} \xrightarrow{\text{endgame}} 1/\|\mu_{t_3}\|_2^2 \sim |G|.$$

Opening. Let t_1 be the maximal integer so that $2t_1 < c_1 \log |G|$ with $c_1 > 0$ from the opening lemma, Lemma 2. Apply the opening lemma to conclude

$$\|\mu_{t_1}\|_2^2 \le \frac{1}{|G|^{\varepsilon_1}}, \tag{3}$$

where $\varepsilon_1 = \varepsilon_1(c_1) > 0$ is a constant. Why does this inequality hold? Well, since the girth of $\mathsf{Cay}(G,S)$ is larger than $2t_1$, a walk of length t_1 on $\mathsf{Cay}(G,S)$ is along a tree. In other words, μ_{t_1} is the probability distribution of a random walk of length t_1 on a $|S|$-ary tree (if S is symmetric). The probability distributions of such random walks are well-understood. Specifically, Kesten [12] proved that

$$\|\mu_{t_1}\|_2^2 \le \left(\frac{2}{|S|}\right)^{t_1},$$

so (3) indeed holds.

Middle-game. We actually need a statistical version of the middle-game lemma, Lemma 3 (see Proposition 2 in [3]). To get a statistical version, apply a version of the Balog-Szemeredi-Gowers theorem (e.g. [21]). We shall not describe the exact statement at this point, but rather hint at the main idea.

The middle-game lemma says that, under some non-triviality assumptions, the size of $A \cdot A \cdot A$ is much larger than that of A. Think of A as the support of μ_{t_1}, and assume that it satisfies the non-triviality assumptions (this requires a proof). The size of A corresponds to $1/\left\|\mu_{t_1}\right\|_2^2$. The set $A \cdot A \cdot A$ corresponds to $\mu_{t_1} * \mu_{t_1} * \mu_{t_1}$. The middle-game lemma says that as long as A is not too large, the set $A \cdot A \cdot A$ is much larger than A. This corresponds to saying that as long as

$$1/\left\|\mu_{t_1}\right\|_2^2 \leq |G|^{1-c_2/2},$$

it holds that

$$1/\left\|\mu_{3t_1}\right\|_2^2 = 1/\left\|\mu_{t_1} * \mu_{t_1} * \mu_{t_1}\right\|_2^2 \geq \left(1/\left\|\mu_{t_1}\right\|\right)^{1+\varepsilon_2}.$$

Here we chose $c_2 > 0$ as the constant from the endgame lemma, Lemma 4, and so $\varepsilon_2 = \varepsilon_2(c_2) > 0$ is a constant. Repeatedly apply this argument for order $1/(\varepsilon_1\varepsilon_2)$ times. In each application, entropy increases by a factor of $1+\varepsilon_2$ in the exponent (as long as it is not already too large). Set t_2 to be roughly $3^{1/(\varepsilon_1\varepsilon_2)}t_1$. Eventually, the norm must be small,

$$\left\|\mu_{t_2}\right\|_2^2 \geq |G|^{1-c_2/2}. \tag{4}$$

Endgame. We now explain the proof for Schreier diagrams (this is different than as hinted above). Apply the endgame lemma with the distribution μ_{t_2} and the second eigenvector f. Since all eigenvectors are orthogonal, and the uniform distribution on X is the first eigenvector, we know $\sum_x f(x) = 0$. The endgame lemma together with (4) imply (set $t_3 = t_2$)

$$\left\|\mu_{t_3} * f\right\|_2^2 \leq \frac{|G|}{|G|^{c_2}} \frac{1}{|G|^{1-c_2/2}} \left\|f\right\|_2^2 = \frac{1}{|G|^{c_2/2}} \left\|f\right\|_2^2,$$

and (2) indeed holds.

So far, the three main steps of the proof, and how to combine them into a proof. In the next sections we discuss each of the three steps in detail.

Remark: It is meant to be the case that each of the three steps/sections that follow can be read independently of the other two.

3 Opening

This is, seemingly, the simplest step of the three. We just need to show that $\mathsf{Cay}(G, S)$ has high girth, that is, that locally it looks like a tree. Finding a single S satisfying this property is, in many cases, indeed quite simple. Proving, however, that *every* S yields (the statistical version of) large girth is much harder, and in general open.

As can be understood from the previous section, even a weaker condition on S is required: That the random walk defined by S behaves similarly to a random walk on a tree, at least for logarithmic time.

We explain the main idea via the example we follow, the graph $\mathsf{Cay}(G, S)$, where $G = \mathsf{SL}_2(\mathbb{F})$ with $\mathbb{F}$ a finite prime field of size p and

$$S = \{g_1, g_2, g_1^{-1}, g_2^{-1}\}$$

with

$$g_1 = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \text{ and } g_2 = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}.$$

The following theorem completes the proof of the opening step in this case, since $|G|$ is order p^3.

Theorem 5. *The girth of* $\mathsf{Cay}(G, S)$ *is at least* $\log_3 p$.

The reader may verify that standard estimates on the number of zeros of polynomials (like the Schwartz-Zippel lemma) in fact imply that the girth of $\mathsf{Cay}(G, S')$ is logarithmic for most choices of S'.

To prove the theorem, translate girth to group terminology. A *reduced* word w is $w = w_1 w_2 \cdots w_t$ with $w_i \in S$ and $w_{i+1} \neq w_i^{-1}$ for all i. The integer t is called the *length* of w. The girth of $\mathsf{Cay}(G, S)$ is at most t iff there is a reduced word w of length at most t so that $w = 1$ in G.

The girth bound (in the theorem above) is proved in two stages. First, we prove that the two matrices g_1, g_2 generate a free group over $\mathbb{Z}$, that is, they do not satisfy any non-trivial relations as integer matrices. In other words, that over $\mathbb{Z}$, every reduced word w is not equal to the identity of $\mathsf{SL}_2(\mathbb{Z})$. Second, we observe that if g_1, g_2 generate a free group over $\mathbb{Z}$, then reducing them modulo p can not yield short relations.

Lemma 6. *The two matrices* g_1, g_2 *generate a free group inside* $\mathsf{SL}_2(\mathbb{Z})$.

The theorem easily follows from the lemma: It follows (by simple induction) that for every integer $j > 0$, the absolute values of the entries of every reduced word of length j in g_1, g_2 in $\mathsf{SL}_2(\mathbb{Z})$ is at most 3^j. The lemma says that in $\mathsf{SL}_2(\mathbb{Z})$ these two elements generate a free group. So, for every $t < \log_3 p$, the reduction modulo p of every reduced word of length t in g_1, g_2 is not identity. The theorem thus holds.

How do one prove that g_1, g_2 generate a free group? One way is a standard geometric argument: the ping-pong lemma (attributed to F. Klein). The general idea is to construct some space X on which g_1, g_2 act. Any non-trivial reduced word w thus defines a map $w(\cdot)$ from X to X. The construction should be such that for every non-trivial w, there is a point x in X so that $w(x) \neq x$. Specifically, $w \neq 1$.

There are many variants of this lemma, but here is a simple one that suffices in our case.

Lemma 7. *Let* g_1, g_2 *act on a set* X. *Assume that there are nonempty disjoint subsets* X_1, X_2 *of* X *so that* $g_1^z(X_2) \subseteq X_1$ *and* $g_2^z(X_1) \subseteq X_2$ *for every nonzero integer* z. *Also assume*[7] *that for every nonzero integer* z, *there exists* $x_0 = x_0(z)$ *in* X *not in* $X_1 \cup X_2$ *so that* $g_1^z(x_0) \in X_1$ *and* $g_2^z(x_0) \in X_2$. *Then,* g_1, g_2 *generate a free group.*

As the name hints, the two elements g_1, g_2 play ping pong between the sets X_1, X_2 and the ball is x_0.

[7] This assumption is not necessary, but suffices for us and makes the proof easier.

Explanation by example. This specific instantiation of the ping pong lemma follows by showing that for every non-trivial reduced word w, we can choose $x_0 \notin X_1 \cup X_2$, so that $w(x_0) \in X_1 \cup X_2$, and specifically $w(x_0) \neq x_0$.

Instead of a formal proof, we give an example that can be easily made into a proof. Consider the word $w = g_1 g_2 g_1^{-2}$. Choose $x_0 = x_0(-2)$, where -2 is the power of the right-most element of w. By assumption, $g_1^{-2}(x_0) \in X_1$. Thus, $g_2(g_1^{-2}(x_0)) \in X_2$ and $w(x_0) = g_1(g_2(g_1^{-2}(x_0))) \in X_1$. □

Using the ping pong lemma, we prove that g_1, g_2 indeed generate a free group.

Proof of Lemma 6. The set X we use is $\mathbb{Z}^2$ on which $G = \mathsf{SL}_2(\mathbb{Z})$ acts by linear transformations. We partition X to

$$X_1 = \{(a,b) : |a| > |b|\} \quad \text{and} \quad X_2 = \{(a,b) : |a| < |b|\}.$$

The sets X_1, X_2 are disjoint and nonempty.

For every nonzero integer z, and $x_2 = (a,b)$ in X_2, we have $g_1^z(x_2) = (a+2zb, b)$. Since $|a| < |b|$, it holds that $|a+2zb| > |b|$, and so $g_1^z(x_2)$ is in X_1. So, $g_1^z(X_2) \subseteq X_1$. Similarly, $g_2^z(X_1) \subseteq X_2$.

It remains to find $x_0 = x_0(z)$ for every given nonzero z. Set $x_0 = (|z|, z)$. We have $g_1^z(x_0) = (|z| + 2z^2, z)$ is in X_1 and $g_2^z(x_0) = (|z|, 2z|z| + z)$ is in X_2. □

4 Middle-game

The middle-game is the most elaborate one. There are much more general versions of the middle game due to Pyber and Szabo [16] and Breuillard, Green and Tao [5]. We focus on the (historically first) approach presented in Bourgain and Gamburd's works [2, 3] that follows Helfgott's work [10]. We explain how to establish the middle-game for the group from our main example $G = \mathsf{SL}_2(\mathbb{F})$, $\mathbb{F}$ a prime field.

Recall that in the middle game we wish to prove product-growth[8]: for $A \subset G$ with certain properties, the size of $A \cdot A \cdot A$ is much larger than that of A. The two properties we require of A is that (i) A generates G and (ii) A is not too large.

The prove consists of six different parts summarized by the following six lemmas. We provide a short intuition for each lemma, but do not provide full proofs. Below we show how the lemmas are combined to a full proof.

Three times suffice. The classical Plunecke-Ruzsa inequality says that if A is a subset of an abelian group G so that $|A \cdot A| \leq C|A|$, then $A_k = \{a_1 a_2 \cdots a_k : a_i \in A\}$ has size $|A_k| \leq C^k |A|$ for every integer k. This statement has been generalized greatly [21], where the reader is referred to for more details.

The following lemma is a special case of the general principle. It is a noncommutative analog of the Plunecke-Ruzsa inequality, that roughly states that if A_3 is not much larger than A, then the same is true for every A_k.

Lemma 8. *There exists a constant $C > 0$ so that the following holds. Let $A \subseteq G = \mathsf{SL}_2(\mathbb{F})$ be so that $|A_3| \leq D|A|$ for some $D > 0$. Then, $|A_k| \leq D^{Ck}|A|$ for every integer k.*

[8] In fact, we wish to prove a statistical version of growth. To move to the statistical language, use the Balog-Szemeredi-Gowers theorem, e.g., as in the work of Tao [21]. We shall not discuss this part of the proof here.

The proof of the lemma uses the notion of approximate groups. It gives a structural characterization of sets A that do not grow under products, such sets are approximate groups. The lemma then easily follows since approximate groups, by definition, do not grow much under products.

Sum-product. This part uses the sum-product theorem to show growth. The sum-product theorem (roughly) states that for every $A \subset \mathbb{F}$, one of the sets $A+A$ or $A \cdot A$ is larger than A. This was used by Helfgott [10] to prove growth in $\mathsf{SL}_2(\mathbb{F})$. The intuition is that when multiplying matrices both sums and products in the underlying field are performed. To move from working with matrices to working with the underlying field, use matrix trace: The *trace* of $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is $\mathsf{Tr} g = a + d$.

Lemma 9. *There exists a constant $c > 0$ so that for every $\delta > 0$, there exists $\varepsilon > 0$ so that the following holds. Let $V \subset \mathsf{SL}_2(\mathbb{F})$ be a set of commuting matrices of size $|V| \le |\mathbb{F}|^{1-\delta}$. Assume that $g \in \mathsf{SL}_2(\mathbb{F})$ is so that, with respect to the basis that makes V diagonal*[9]*, $g_{1,1}g_{1,2}g_{2,1}g_{2,2} \neq 0$. Then,*

$$|\mathsf{Tr} V_8 g V_8 g^{-1}| \ge c|V|^{1+\varepsilon}.$$

The lemma follows by reduction to the sum-product theorem, using *Rusza distances.* Here is a very rough outline. Diagonal matrices in $\mathsf{SL}_2(\mathbb{F})$ have the form $v = \begin{pmatrix} x & 0 \\ 0 & 1/x \end{pmatrix}$. For two diagonal matrices v_1, v_2, the matrix $v_1 g v_2 g^{-1}$ "contains" a map of the form $x \mapsto x + 1/x$. The proof boils down to analyzing the growth that this map induces.

Finding commutative set. In light of the previous lemma, we need to find a commutative set of matrices. This part uses matrix-trace to find a "large" commuting set of matrices.

Lemma 10. *For every $A \subseteq \mathsf{SL}_2(\mathbb{F})$ so that*[10] *$A = A^{-1}$, there exists a set $V \subseteq A_2$ of commuting matrices so that*

$$|V| \ge |\mathsf{Tr} A| \frac{|A|}{|A_3|}.$$

The proof of the lemma uses that trace gives knowledge of eigenvalues. To prove the lemma, choose the least common trace in A_3 and use its pre-image under trace to find V.

Trace-set is large. To obtain a useful bound on the size of V from the previous lemma, we require that $\mathsf{Tr} A$ is large.

Lemma 11. *There exist constants $c, k > 0$ so that the following holds. For every $A \subseteq \mathsf{SL}_2(\mathbb{F})$ so that $\langle A \rangle = \mathsf{SL}_2(\mathbb{F})$,*

$$|\mathsf{Tr} A_k| \ge c|A|^{1/3}.$$

This part uses that trace is a "linear projection" of matrices to show that the trace of a set is large. The group $\mathsf{SL}_2(\mathbb{F})$ is a three-dimensional object, so a set of size s in it, should have a linear projection of size at least $s^{1/3}$, as the lemma says.

[9] A set of commuting matrices can be simultaneously diagonalized.

[10] That A is symmetric is not essential but makes the formulation simpler.

Algebraic growth. The previous lemmas mostly dealt with V, a commuting set of matrices, which is, therefore, of size at most $|\mathbb{F}|$. This is too small in comparison to the size of A, which can be of order, say, $|\mathbb{F}|^2$. This part uses simple properties of polynomials to move back to the "correct" order of magnitude.

Lemma 12. *There exists a constant $c > 0$ so that the following holds. Let $V \subset \mathsf{SL}_2(\mathbb{F})$ be a set of commuting matrices of size $|V| \geq 30$. Assume that $g \in \mathsf{SL}_2(\mathbb{F})$ is so that, with respect to the basis that makes V diagonal, $g_{1,1}g_{1,2}g_{2,1}g_{2,2} \neq 0$. Then,*

$$|VgVg^{-1}V| \geq c|V|^3.$$

The lemma (roughly) follows by showing that the map

$$(x, y, z) \mapsto \begin{pmatrix} x & 0 \\ 0 & 1/x \end{pmatrix} g \begin{pmatrix} y & 0 \\ 0 & 1/y \end{pmatrix} g^{-1} \begin{pmatrix} z & 0 \\ 0 & 1/z \end{pmatrix}$$

is finite, that is, the pre-image of (most) image-points is of finite size.

Generation yields useful elements. Two of the lemmas above use a "useful" element g, i.e., so that after a basis change $g_{1,1}g_{1,2}g_{2,1}g_{2,2} \neq 0$. We need to find a useful element inside A or some iterated product of it.

Lemma 13. *There exists an integer k so that the following holds. Let $A \subseteq \mathsf{SL}_2(\mathbb{F})$, $|\mathbb{F}| > 3$, be so that $\langle A \rangle = \mathsf{SL}_2(\mathbb{F})$. Then, with respect to any basis change in*[11] *$\mathsf{SL}_2(\bar{\mathbb{F}})$, there is $g \in A_k$ so that $g_{1,1}g_{1,2}g_{2,1}g_{2,2} \neq 0$.*

The existence of a useful g follows since A generates the whole group. For example, the set of upper triangular matrices (which do not satisfy $g_{1,1}g_{1,2}g_{2,1}g_{2,2} \neq 0$) is a subgroup, but by assumption A contains an element outside of it.

The lemmas above are combined in a simple but clever way to prove product growth, Lemma 3. To simplify notation, in the following c, k are universal constants that may change their value from time to time, $\delta > 0$ is a fixed constant, and $\varepsilon = \varepsilon(\delta) > 0$ may change its value as well.

Assume, for simplicity, that A is symmetric, $A^{-1} = A$. In light of Lemma 8 it suffices to prove that $|A_{k_0}| \geq c|A|^{1+\varepsilon_0}$ for some constants k_0, ε_0. This is what we shall do. Assume towards a contradiction that

$$|A_{k_0}| < c|A|^{1+\varepsilon_0}. \tag{5}$$

First, use Lemma 10 with the set A_k, where k is the constant from Lemma 11. We thus found a commutative set of matrices $V \subset A_{2k}$ so that

$$|V| \geq |\mathsf{Tr}A_k| \frac{|A_k|}{|A_{3k}|}.$$

Lemma 11 and assumption (5) imply

$$|V| \geq c|A|^{1/3} \frac{1}{|A|^{\varepsilon_0/2}} \geq c|A|^{1/3-\varepsilon_0/2}.$$

[11] $\bar{\mathbb{F}}$ is the algebraic closure of $\mathbb{F}$. Algebraic closure is necessary to diagonalize all matrices.

Second, use Lemma 10 with the set $U = V_8 g V_8 g^{-1} V_8$, where g is the useful element given by Lemma 13 w.r.t. the basis that makes V_8 diagonal. We found a commutative set of matrices $W \subset U_2$ so that

$$|W| \geq |\mathsf{Tr}U| \frac{|U|}{|U_3|}.$$

By Lemma 9 and choice of g, we know $|\mathsf{Tr}U| \geq c|V|^{1+\varepsilon_1}$, where $\varepsilon_1 > 0$ is a constant. Using Lemma 12 and assumption (5), since $U_3 \subset A_k$,

$$|W| \geq |\mathsf{Tr}U| \frac{|U|}{|U_3|} \geq c|V|^{1+\varepsilon_1} \frac{|V|^3}{|A_k|} \geq c|A|^{1/3+\varepsilon}.$$

Finally, let g be a useful element w.r.t. the basis that makes W diagonal as given by Lemma 12,

$$|A_{k_0}| \geq |WgWg^{-1}W| \geq c|W|^3 \geq c|A|^{1+\varepsilon_0}.$$

This contradicts (5).

5 Endgame

The endgame is about obtaining a non-trivial estimate on the behavior of convolution. There are two different approaches for establishing this. One is using representation theory: Sarnak and Xue's multiplicity argument [19] or Gower's notion of *quasi-random group* [9]. The other is using two-transitivity as in [4]. We shall explain both approaches. The first approach we consider is using two-transitivity, since the notions it uses are simpler.

5.1 Two-transitivity

We now explain how do two-transitive actions imply the endgame. This approach is useful for a special type of Schreier diagrams: when G's action on X is two-transitive (or pairwise independent). This approach towards establishing the endgame was used in [4] where a monotone expander was constructed. It is especially relevant to cases when G is not compact, since then the quasi-random groups approach that is based on representation theory does not work.

Lemma 14. *Let G be a finite group that acts two-transitively on a finite set X. Let μ be a probability distribution on G. Let $f : X \to \mathbb{R}$ be so that $\sum_{x \in X} f(x) = 0$. Then,*

$$\|\mu * f\|_2^2 \leq \frac{|G|}{(|X|-1)^{1/2}} \|\mu\|_2^2 \|f\|_2^2 .$$

The lemma implies the mixing property, Lemma 4, as long as $|X|$ is polynomially comparable to $|G|$. For the example we follow, the Schreier graph defined by the Möbius action, the lemma completes the endgame, since the action is indeed two-transitive and $|X|$ is order $|G|^{1/3}$.

The proof of the lemma is just a few lines of calculation that can be summarized as: first apply Cauchy-Schwartz to remove μ, and then use two-transitivity in a straightforward way to obtain non-trivial cancellations. Where does the "$|G|/|X|$ factor" come from? Well, in a nutshell, from the following equality: for every $f : X \to \mathbb{R}$ and for every x in X, it holds that $\sum_{g \in G} f(g(x))^2 = (|G|/|X|) \|f\|_2^2$.

Proof. First, use Cauchy-Schwartz inequality to remove the dependence on μ,

$$\begin{aligned}
\|\mu * f\|_2^2 &= \sum_{g,g'} \mu(g)\mu(g') \sum_x f(g^{-1}(x)) f({g'}^{-1}(x)) \\
&\le \|\mu\|_2^2 \left(\sum_{g,g'} \sum_{x,x'} f(g^{-1}(x)) f(g^{-1}(x')) f({g'}^{-1}(x)) f({g'}^{-1}(x')) \right)^{1/2} \\
&= \|\mu\|_2^2 \left(\sum_{x,x'} \left(\sum_g f(g^{-1}(x)) f(g^{-1}(x')) \right)^2 \right)^{1/2} = \dots
\end{aligned}$$

Now, for fixed $x \neq x'$, since $\sum_x f(x) = 0$ and since the action is two-transitive,

$$\begin{aligned}
\sum_g f(g^{-1}(x)) f(g^{-1}(x')) &= \sum_{x''} f(x'') \sum_{g : g^{-1}(x) = x''} f(g^{-1}(x')) \\
&= \sum_{x''} f(x'') \sum_{x''' \neq x''} \frac{|G|}{|X|(|X|-1)} f(x''') \\
&= - \|f\|_2^2 \frac{|G|}{|X|(|X|-1)}.
\end{aligned}$$

So, we can continue

$$\begin{aligned}
\dots &= \|\mu\|_2^2 \, \|f\|_2^2 \left(|X| \frac{|G|^2}{|X|^2} + |X|(|X|-1) \frac{|G|^2}{|X|^2(|X|-1)^2} \right)^{1/2} \\
&= \|\mu\|_2^2 \, \|f\|_2^2 \, |G| / (|X|-1)^{1/2}.
\end{aligned}$$

□

5.2 Multiplicity and quasirandomness

The second approach towards establishing a mixing property is via representation theory: Recall that we just wish to establish a bound on the eigenvalues of a given matrix M (that is defined by the graph in question). It would have been extremely simple to do so, if the matrix was diagonal. This is, of course, too good to be true. Since M is defined by a group action, we can almost diagonalize it using representation theory, as we now explain.

We start with a brief introduction of basic concepts (for more details, see Serre's book [20]). Let G be a finite group. We can look for "copies" of G inside matrix groups, specifically, inside the group of $m \times m$ invertible complex matrices $\mathsf{GL}_m(\mathbb{C})$. A *representation*[12] is a map $\rho : G \to \mathsf{GL}_m(\mathbb{C})$ that respects the group operation, that is, for every g, g' in G, it holds that $\rho(g)\rho(g') = \rho(gg')$, as matrices. The integer m is called the *dimension* of ρ.

Two standard examples: The *trivial* representation maps every g in G to the identity matrix. The *regular* representation captures the action of G on itself; it is a map ρ from G to $\mathsf{GL}_{|G|}(\mathbb{C})$

[12] We shall only consider representations over $\mathbb{C}$. For concreteness, we only consider matrix-representations (instead of the linear transformations they define).

defined by: for every g in G, the matrix $\rho(g)$ is

$$(\rho(g))_{g_1,g_2} = \begin{cases} 1 & gg_1 = g_2, \\ 0 & gg_1 \neq g_2. \end{cases}$$

Similarly to that integers have building blocks that are called prime numbers, representations have building blocks that are called irreducible representations. A representation ρ is *irreducible* if every G-invariant subspace of $\mathbb{C}^m$ is trivial, that is, if U is a subspace of $\mathbb{C}^m$ so that $g(U) \subset U$ for every g in G, then U is either $\mathsf{GL}_m(\mathbb{C})$ or $\{0\}$. If, e.g., ρ is irreducible and trivial, then it is one-dimensional and $\rho(g) = 1$ for all g.

There is obviously an infinite number of irreducible representation since there are infinitely many choices of basis. We shall thus say that two representations ρ, ρ' are *isomorphic* if there is a matrix s (a basis change) so that $\rho' = s\rho s^{-1}$.

Theorem 15. *There is a finite list $\rho_1, \ldots, \rho_t$ of (non-isomorphic) irreducible representation so that every representation ρ of G can be written as a direct sum of copies of $\rho_1, \ldots, \rho_t$.*

The theorem says, in other words, that for every representation ρ of G, w.r.t. some choice of basis, ρ is in block-diagonal form with blocks from the list $\rho_1, \ldots, \rho_t$.

Representations allow to state a sufficient condition under which the mixing property holds. This condition was considered in several works, and Gowers calls it quasi-randomness [9]. A group G is *D-quasi-random* if every non-trivial irreducible representation of G has dimension at least D. The following theorem proved by Babai, Nikolov and Pyber [1] summarizes the statement.

Theorem 16. *Assume G is a D-quasi-random group that acts transitively on X. Let $\mu : G \to \mathbb{R}$ and $f : X \to \mathbb{R}$ be so that $\sum_{x \in X} f(x) = 0$. Then,*

$$\|\mu * f\|_2^2 \leq \frac{|G|}{D} \|\mu\|_2^2 \|f\|_2^2 .$$

The theorem shows that a mixing property holds for any group G that is D-quasi-random with D that is polynomially comparable to $|G|$. To prove the theorem[13] we shall use properties of the regular representation (see e.g. [20]) as was done in [19].

Proof of Theorem 16. We start the proof by stating some known properties of the regular representation. Let $\rho_1, \ldots, \rho_t$ be the list of all (non-isomorphic) irreducible representations of G, where ρ_1 is the trivial representation. Let ρ be the regular representation of G. Theorem 15 implies that ρ can be written as a direct sum of $\rho_1, \ldots, \rho_t$. Denote by c_i the number of copies of ρ_i in ρ. Denote by d_i the dimension of ρ_i.

Theorem 17 (structure of regular representation)**.** *$c_i = d_i$ for every $i \in [t]$.*

We use these properties of the regular representation to bound the relevant singular values. First, some definitions. For an $k \times k$ complex matrix M, denote by M^* the conjugate transpose of M. The matrix MM^* is positive semi-definite, and has k non-negative eigenvalues $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_k \geq 0$. Denote

$$\sigma(M) = \sigma_1.$$

[13] A different way to prove the theorem is using Schur's orthogonality (see [20]).

Denote

$$M_i = \sum_{g \in G} \mu(g)\rho_i(g)$$

for $1 \leq i \leq t$. By Theorem 17, after a basis change, the matrix

$$R = \sum_{g \in G} \mu(g)\rho(g)$$

is in block-diagonal form with blocks from the list $M_1, \ldots, M_t$, each M_i has multiplicity c_i. Denote by Q the $|G| - 1 \times |G| - 1$ matrix with blocks of the form $M_2, \ldots, M_t$, after deleting from R a row and a column that correspond to the trivial representation ($c_1 = d_1 = 1$). On one hand, quasirandomness implies

$$\mathrm{trace}(QQ^*) = \sum_{2 \leq i \leq t} c_i \mathrm{trace}(M_i M_i^*) \geq D \cdot \sigma(Q).$$

On the other hand,

$$\mathrm{trace}(QQ^*) \leq \mathrm{trace}(RR^*) = \sum_{g_1 \in G} \sum_{g_2 \in G} \mu(g_2 g_1^{-1})^2 = |G| \, \|\mu\|_2^2 \, .$$

We conclude that

$$\max\{\sigma(M_i) : 2 \leq i \leq t\} = \sigma(Q) \leq \|\mu\|_2^2 \, |G|/D.$$

Finally, we use the bound on singular values to prove the theorem. Consider the representation ρ' of G induced by the action of G on X: for every g in G, the $|X| \times |X|$ matrix $\rho'(g)$ is defined by the linear map $(\rho'(g)h)(x) = h(g^{-1}x)$ for $h : X \to \mathbb{C}$. By Theorem 15, after a basis change, the matrix $R' = \sum_{g \in G} \mu(g)\rho'(g)$ can be written in block-diagonal form with blocks that are copies of $M_1, \ldots, M_t$, each of the blocks appears in R' with multiplicity c'_i.

We claim that $c'_1 = 1$, since G acts (one-) transitively on X. There is a formula (that follows from Schur's orthogonality relations, see [20]) for calculating c'_i, and specifically

$$c'_1 = \frac{1}{|G|} \sum_{g \in G} \mathrm{trace}(\rho'(g)) = \frac{1}{|G|} \sum_{g \in G} \sum_{x \in X} \rho'(g)_{x,x} = \frac{1}{|G|} \sum_{x \in X} \big|\{g \in G : g(x) = x\}\big| = 1.$$

The last equality holds due to transitivity, which implies that $\big|\{g \in G : g(x) = x\}\big| = \big|\{g \in G : g(x) = x'\}\big|$ for every x, x' in X.

Since $\langle f, \bar{1} \rangle = \sum_x f(x) = 0$ and since the all-ones vector spans the subspace corresponding to the trivial representation,

$$\|\mu * f\|_2^2 = \langle R' f, R' f \rangle \leq \sigma(Q) \, \|f\|_2^2 \, .$$

□

Acknowledgments

I wish to thank Peter Sarnak and Avi Wigderson for very helpful comments.

References

[1] L. Babai, N. Nikolov and L. Pyber, *Product growth and mixing in finite groups*, SODA, pages 248–257, 2008.

[2] J. Bourgain and A. Gamburd, *On the spectral gap for finitely-generated subgroups of SU(2)*, Inventiones Mathematicae 171 (1), pages 83–121, 2007.

[3] J. Bourgain and A. Gamburd, *Uniform expansion bounds for Cayley graphs of* $SL_2(F_p)$, Annals of Mathematics 167 (2), pages 625–642, 2008.

[4] J. Bourgain and A. Yehudayoff, *Monotone expansion*, STOC, pages 1061–1078, 2012.

[5] E. Breuillard, B. Green and T. Tao, *The structure of approximate groups,* arXiv:1110.5008, 2011.

[6] G. Davidoff, P. Sarnak and A. Valette, *Elementary number theory, group theory and Ramanujan graphs,* London Mathematical Society Student Texts (No. 55).

[7] C. M. Dawson and M. A. Nielsen, *The Solovay–Kitaev algorithm,* Quantum Inf. Comput. 6, pages 81–95, 2006.

[8] O. Gabber and Z. Galil, *Explicit constructions of linear-sized superconcentrators,* J. Comput. System Sci. 22 (3), pages 407-420, 1981. Special issue dedicated to Michael Machtey.

[9] W. T. Gowers, *Quasirandom groups,* Combinatorics, Probability and Computing 17, pages 363–387, 2008.

[10] H. A. Helfgott, *Growth and generation in* $SL_2(Z/pZ)$, Annals of Mathematics 167 (2), pages 601–623, 2008.

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

[12] H. Kesten, *Symmetric random walks on groups,* Trans. Am. Math. Soc. 92, pages 336–354, 1959.

[13] A. Lubotzky, *Cayley graphs: eigenvalues, expanders and random walks,* Surveys in Combinatorics (P. Rowlinson ed.), London Math. Soc. Lecture Note Ser. 18, pages 155–189, Cambridge Univ. Press, Cambridge, 1995.

[14] A. Lubotzky, R. Phillips and P. Sarnak, *Ramanujan graphs,* Combinatorica 8, pages 261–277, 1988.

[15] G. A. Margulis, *Explicit constructions of expanders*, Problemy Peredaci Informacii 9 (4), pages 71–80, 1973.

[16] L. Pyber and E. Szabo, *Growth in finite simple groups of Lie type of bounded rank,* arXiv:1005.1858, 2010.

[17] O. Reingold, S. Vadhan and A. Wigderson, *Entropy waves, the zig-zag graph product, and new constant-degree expanders*, Annals of Mathematics 155 (1), pages 157–187, 2002.

[18] P. Sarnak, *Notes on thin groups,* Notes prepared for MSRI hot topics workshop on super-strong approximations, 2012.

[19] P. Sarnak and X. Xue, *Bounds for multiplicities of automorphic representations.* Duke Math. J. 64, pages 207–227, 1991.

[20] J. P. Serre, *Linear representations of finite groups.* Springer-Verlag New York Inc., 1977.

[21] T. Tao, *Product set estimates for non-commutative groups,* Combinatorica 28 (5), pages 547–594, 2008.