EXPLICIT GROWTH AND EXPANSION FOR SL_2

EMMANUEL KOWALSKI

Abstract. We give explicit versions of Helfgott's Growth Theorem for SL_2, as well as of the Bourgain-Gamburd argument for expansion of Cayley graphs modulo primes of subgroups of $\mathrm{SL}_2(\mathbf{Z})$ which are Zariski-dense in SL_2.

Contents

1. Introduction

Our main goal in this paper is to prove the following result, which is an explicit version of a theorem of Bourgain and Gamburd [1]:

Theorem 1.1. *Let $S\subset \mathrm{SL}_2(\mathbf{Z})$ be a finite symmetric set such that the subgroup generated by S is Zariski-dense in $\mathrm{SL}_2(\mathbf{Z})$. Let $\mathcal{P}$ be the set of primes such that $S_p = S \,(\mathrm{mod}\, p)$ generates $\mathrm{SL}_2(\mathbf{F}_p)$, which contains all but finitely many primes. Then the family of Cayley graphs $(\mathcal{C}(\mathrm{SL}_2(\mathbf{F}_p), S_p))_{p\in\mathcal{P}}$ is an expander family, and one can write down explicit bounds for the spectral gap, given the set S.*

1991 *Mathematics Subject Classification.* 20F69, 05C50, 05C81.

Key words and phrases. Growth of finite groups, expander graphs, Cayley graphs, diameter, random walks on groups.

In particular, if S generates a free group of rank $|S|/2$, the spectral gap[1] *satisfies*

$$\lambda_1(\mathcal{C}(\mathrm{SL}_2(\mathbf{F}_p), S_p)) \geqslant 1 - \exp\Big(-\frac{6}{j}\Big), \tag{1.1}$$

for all p large enough, where[2]

$$j = 20000 \max(1008\gamma^{-1}, 1)$$

and

$$\gamma = \frac{1}{2^{10}} \frac{\log |S|}{\log \max_{s\in S} \|s\|},$$

the norm $\|s\|$ being the operator norm of the matrix s, with respect to the euclidean metric on $\mathbf{C}^2$.

We can specify what "p large enough" means, but we defer a statement to Section 4.3 since this involves a series of inequalities which are awkward to state (and unenlightening), but easy to check for a given concrete set of matrices S.

A crucial ingredient in the argument of Bourgain and Gamburd is Helfgott's Growth Theorem [7] for SL_2, which has considerable independent interest. We thus require an explicit version of it, and we will prove the following:

Theorem 1.2. *Let p be a prime number, $H \subset \mathrm{SL}_2(\mathbf{F}_p)$ a symmetric generating subset of $\mathrm{SL}_2(\mathbf{F}_p)$ containing 1. Then the triple product set $H^{(3)} = H \cdot H \cdot H$ satisfies either $H^{(3)} = \mathrm{SL}_2(\mathbf{F}_p)$ or*

$$|H^{(3)}| \geqslant |H|^{1+\delta},$$

where $\delta = 1/1344$.

Here is a simple corollary, which is (as far as the author is aware) also the first explicit result of this kind for almost simple linear groups:

Corollary 1.3 (Explicit solution to Babai's conjecture for $\mathrm{SL}_2(\mathbf{F}_p)$)**.** *For any prime number p and any symmetric generating set S of $\mathrm{SL}_2(\mathbf{F}_p)$, we have*

$$\mathrm{diam}\, \mathcal{C}(\mathrm{SL}_2(\mathbf{F}_p), S) \leqslant 3(\log |\mathrm{SL}_2(\mathbf{F}_p)|)^C$$

with $C = 1345$.

Another corollary of Helfgott's Theorem and of intermediate results used in the proof of Theorem 1.1 is a better diameter bound for Zariski-dense subgroups:

Corollary 1.4 (Diameter bounds for Zariski-dense subgroups of SL_2)**.** *Let $S \subset \mathrm{SL}_2(\mathbf{Z})$ be a finite symmetric set such that the subgroup generated by S is Zariski-dense in $\mathrm{SL}_2(\mathbf{Z})$ and is a free group of rank $|S|/2$. Let $\mathcal{P}$ be the set of primes such that $S_p = S \,(\mathrm{mod}\, p)$ generates $\mathrm{SL}_2(\mathbf{F}_p)$.*

Let $\delta > 0$ be as in Helfgott's Theorem and define

$$\tau^{-1} = \log \max_{s\in S} \|s\| > 0.$$

Then for $p \in \mathcal{P}$ and $p > \exp(2/\tau)$, we have

$$\mathrm{diam}(\mathcal{C}(\mathrm{SL}_2(\mathbf{F}_p), S)) \leqslant 3^A (\log |\mathrm{SL}_2(\mathbf{F}_p)|)$$

[1] This is the spectral gap of the *normalized* Laplace operator $\Delta = \mathrm{Id} - M$, where M is the Markov averaging operator of the graph; thus the spectrum of Δ is a subset of the interval $[0, 2]$.

[2] The value of j is certainly almost always, if not always, the one given by the first alternative.

where

$$A = \frac{\log(8\tau^{-1}(|S|-1)^{-1})}{\log(1+\delta)}.$$

Remark 1.5. Using the well-known bound

$$\lambda_1(\Gamma_p) \geqslant \frac{1}{|S|\,\mathrm{diam}(\Gamma_p)^2}$$

(see, e.g., [13, Th. 13.23]), these diameter bounds can be used to get lower bounds for spectral gaps for "medium" primes. Note the huge discrepancy however at the end of the range.

Combining Theorem 1.1 with the second corollary, we can give explicit statements for the motivating example of the Lubotzky group.

Corollary 1.6 (The Lubotzky group)**.** *Let*

$$S = \left\{ \begin{pmatrix} 1 & \pm 3 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ \pm 3 & 1 \end{pmatrix} \right\} \subset \mathrm{SL}_2(\mathbf{Z}),$$

and let $\Gamma_p = \mathcal{C}(\mathrm{SL}_2(\mathbf{F}_p), S_p)$. *Then we have*

$$\lambda_1(\Gamma_p) \geqslant 2^{-32} \tag{1.2}$$

if $p \geqslant 2^{2^{47}}$, *and*

$$\mathrm{diam}(\Gamma_p) \leqslant 2^{2478}(\log|\,\mathrm{SL}_2(\mathbf{F}_p)|)$$

for all $p \neq 3$.

Remark 1.7. The gap 2^{-32} is not very large, but is not ridiculously small either. One can accurately describe it as "astronomically" small: if one were to think of the interval $[0,2]$ as a straight path from Neptune to the Sun,[3] the spectral gap (1.2) says roughly that we can be sure that no eigenvalue of the Laplace operator for the Cayley graphs above (for $p \geqslant 2^{2^{47}}$) ever occurs during the first kilometer.

The original papers of Bourgain and Gamburd [1] and Helfgott [7] are effective, and thus it is not surprising that one can obtain explicit versions. What is less clear is how good the constants may be, and how much work may be required to provide them. This paper gives a first indication in that respect. It is maybe interesting to note that small tweaks to the original proof (especially in the context of the "L^2-flattening lemma" of [1]) can have big consequences on the size of the spectral gap.

The bounds we derive are very unlikely to be anywhere near sharp, and not only because we often use rather rough estimates to simplify the shape and constants appearing in various inequalities.[4] Indeed, when the Hausdorff dimension of the limit set of the subgroup G generated by S is large enough, Gamburd [4] has shown quite good spectral gaps for the hyperbolic Laplace operator on $G\backslash\mathbf{H}$, which strongly suggest that the corresponding combinatorial spectral gap would be also relatively large. But this computation has not been done, to the author's knowledge, and our version of Theorem 1.1 gives the first fully explicit spectral gap for infinite-index subgroups of $\mathrm{SL}_2(\mathbf{Z})$, with Corollary 1.6 being a nice concrete example (it is also known that the "Lubotzky group" is too small for Gamburd's result to apply).

[3] About 4,452,940,833 kilometers at perihelion.

[4] In some cases, one can easily extract better bounds from the proof, e.g., one can replace 1/1344 by 1/673 for all H large enough in Theorem 1.2.

In view of the direct link between the spectral gap of families of Cayley graphs of quotients of "thin" (or sparse) subgroups of arithmetic groups and quantitative applications of sieve methods to these groups, it is natural to wish for a better understanding of these issues.[5] A first step towards effective versions of these applications of "sieve in orbit" would be to extend Theorem 1.1 to an effective spectral gap for $\mathrm{SL}_2(\mathbf{Z}/q\mathbf{Z})$, where q is a squarefree modulus (as originally proved by Bourgain, Gamburd and Sarnak [2]), and we hope to come back to this.

As a final remark, the reader can also see this paper as presenting a complete proof of the qualitative forms of Theorems 1.1 and 1.2 and their corollaries. When read in this light, ignoring the fussy technical details arising from trying to have explicit bounds, it may in fact be useful as a self-contained introduction to this area of research.

Notation. As usual, $|X|$ denotes the cardinality of a set. Given a group G, and a symmetric generating set S, we denote by $\mathcal{C}(G, S)$ the Cayley graph of G with respect to S, which is $|S|$-regular. Moreover, we say that a symmetric set $S \subset G$ *freely generates* G if representatives of S modulo the relation $s \sim s^{-1}$ form a free generating set of G, i.e., G is a free group of rank $|S|/2$.

For a subset $H \subset G$ of a group G, we write $H^{(n)}$ for the n-fold product set

$$H^{(n)} = \{x \in G \mid x = h_1 \cdots h_n \text{ for some } h_i \in H\}.$$

Note the immediate relations

$$(H^{(n)})^{(m)} = H^{(nm)}, \quad H^{(n+m)} = H^{(n)} \cdot H^{(m)}$$

for n, $m \geqslant 0$ and $(H^{(n)})^{-1} = H^{(n)}$ if H is symmetric. In addition, if $1 \in H$, we have $H^{(n)} \subset H^{(m)}$ for all $m \geqslant n$. In particular, the diameter of a Cayley graph $\mathcal{C}(G, H)$, when $H = H^{-1}$, is the smallest $n \geqslant 1$ such that $\tilde{H}^{(n)} = G$, where $\tilde{H} = H \cup \{1\}$.

We denote by $\mathrm{trp}(H)$ the "tripling constant" of a subset $H \subset G$, defined by

$$\mathrm{trp}(H) = \frac{|H^{(3)}|}{|H|}.$$

Acknowledgments. Much of the work on this paper was done during and following a course on expander graphs that I taught at ETH Zürich during the Fall Semester 2011.[6] Thanks to all who attended this course and helped with corrections and remarks. Thanks in particular to O. Dinai, and to R. Pink for very interesting discussions and for helping with the proof of the specific variant of a "Larsen-Pink" inequality in Theorem 3.10. Thanks very much to P. Sarnak for his comments, and especially for his insights concerning the history of questions and results concerning spectral gaps for subgroups of $\mathrm{SL}_2(\mathbf{Z})$. Thanks to L. Pyber for clarifying some of the "combinatorics" in the proof of the growth theorem.

2. Explicit multiplicative combinatorics

Another ingredient of Theorem 1.1 is the relation between subsets of a finite group with small "multiplicative energy" and sets with small tripling constant, or approximate subgroups. This was obtained by Tao [18], in good qualitative form, but without explicit dependency of the various quantities involved. In this section, we state a suitably explicit version.

[5] Indeed, this question was asked by J-P. Serre during the author's Bourbaki lecture [10].

[6] Lecture notes for this course are available, and contain more background and motivating material [11].

We recall first the definitions involved. For a finite group G and A, $B \subset G$, one defines the multiplicative energy by

$$E(A,B) = |\{(a_1,a_2,b_1,b_2) \in A^2 \times B^2 \mid a_1b_1 = a_2b_2\}|.$$

It is also convenient to denote by

$$e(A,B) = \frac{|E(A,B)|}{(|A||B|)^{3/2}}.$$

the normalized multiplicative energy, which is $\leqslant 1$. Following Tao (see [18, Def. 3.8]), for a finite group G and any $\alpha \geqslant 1$, a subset $\mathsf{H} \subset G$ is an α-approximate subgroup if $1 \in \mathsf{H}$, $\mathsf{H} = \mathsf{H}^{-1}$ and there exists a subset $X \subset G$ of order at most α such that

$$\mathsf{H} \cdot \mathsf{H} \subset X \cdot \mathsf{H}, \tag{2.1}$$

which implies also $\mathsf{H} \cdot \mathsf{H} \subset \mathsf{H} \cdot X$. Then we have:

Theorem 2.1. *Let G be a finite group and $\alpha \geqslant 1$. If A and B are subsets of G such that $e(A,B) \geqslant \alpha^{-1}$, there exist constants β_1, β_2, $\beta_3 \geqslant 1$, a β_1-approximate subgroup $\mathsf{H} \subset G$ and elements x, $y \in G$ such that*

$$|\mathsf{H}| \leqslant \beta_2|A| \leqslant \beta_2\alpha^2|B|,$$
$$|A \cap x\mathsf{H}| \geqslant \frac{1}{\beta_3}|A|, \qquad |B \cap \mathsf{H}y| \geqslant \frac{1}{\beta_3}|B|,$$
$$\mathrm{trp}(\mathsf{H}) \leqslant \beta_4,$$

and moreover $\beta_i \leqslant c_1\alpha^{c_2}$ for some absolute *constants c_1, $c_2 > 0$. In fact, one can take*

$$\beta_1 \leqslant 2^{1621}\alpha^{720}, \qquad \beta_2 \leqslant 2^{283}\alpha^{126}, \qquad \beta_3 \leqslant 2^{2112}\alpha^{937}, \qquad \beta_4 \leqslant 2^{810}\alpha^{360}. \tag{2.2}$$

Except for the values of the constants, this is proved in [18, Th. 5.4, (i) implies (iv)] and quoted in [19, Th. 2.48]. Since this is obtained by following line by line the arguments of Tao, we defer a proof to the Appendix.

3. Growth for SL_2

We prove here Theorem 1.2. The argument we use is basically the one sketched by Pyber and Szabó in [16, §1.1] (which is expanded in their paper to cover much more general situations). It is closely related to the one of Breuillard, Green and Tao [3], and many ingredients are already visible in Helfgott's original argument [7].

3.1. Elementary facts and definitions.

We begin with an important observation, which applies to all finite groups, and goes back to Ruzsa: to prove that the tripling constant of a generating set H is at least a small power of $|H|$, it is enough to prove that the growth ratio after an arbitrary (but fixed) number of products is of such order of magnitude.

Proposition 3.1 (Ruzsa)**.** *Let G be a finite group, and let $H \subset G$ be a symmetric non-empty subset.*

(1) *Denoting $\alpha_n = |H^{(n)}|/|H|$, we have*

$$\alpha_n \leqslant \alpha_3^{n-2} = \mathrm{trp}(H)^{n-2} \tag{3.1}$$

for all $n \geqslant 3$.

(2) *We have $\mathrm{trp}(H^{(2)}) \leqslant \mathrm{trp}(H)^4$ and for $k \geqslant 3$, we have*

$$\mathrm{trp}(H^{(k)}) \leqslant \mathrm{trp}(H)^{3k-3}.$$

Proof. The first part is well-known. For (2), we have

$$\mathrm{trp}(H^{(k)}) = \frac{h_{3k}}{h_k} = \frac{\alpha_{3k}}{\alpha_k}.$$

Since $\alpha_k \geqslant \alpha_3$ for $k \geqslant 3$, we obtain $\mathrm{trp}(H^{(k)}) \leqslant \alpha_3^{3k-3}$ for $k \geqslant 3$ by (1), while for $k \geqslant 2$, we simply use $\alpha_2 \geqslant 1$ to get $\mathrm{trp}(H^{(2)}) \leqslant \alpha_3^4$. □

We first use of Ruzsa's Lemma to show that Helfgott's Theorem holds when $|H|$ is small, in the following sense:

Lemma 3.2. *Let G be a finite group and let H be a symmetric generating set of G containing* 1*. If $H^{(3)} \neq G$, we have $|H^{(3)}| \geqslant 2^{1/2}|H|$.*

Proof. If the triple product set is not all of G, it follows that $H^{(3)} \neq H^{(2)}$. We fix some $x \in H^{(3)} - H^{(2)}$, and consider the injective map

$$i : \begin{cases} H & \longrightarrow & G \\ h & \mapsto & hx \end{cases}.$$

The image of this map is contained in $H^{(4)}$ and it is disjoint with H since $x \notin H^{(2)}$. Hence $H^{(4)}$, which contains H and the image of i, satisfies $|H^{(4)}| \geqslant 2|H|$. Hence, by Ruzsa's Lemma, we obtain

$$\mathrm{trp}(H) \geqslant \Big(\frac{|H^{(4)}|}{|H|}\Big)^{1/2} \geqslant 2^{1/2}.$$

□

The following version of the orbit-stabilizer theorem will be used to reduce the proof of lower-bounds on the size a set to an upper-bound for another.

Proposition 3.3 (Helfgott)**.** *Let G be a finite group acting on a non-empty finite set X. Fix some $x \in X$ and let $K \subset G$ be the stabilizer of x in G. For any non-empty symmetric subset $H \subset G$, we have*

$$|K \cap H^{(2)}| \geqslant \frac{|H|}{|H \cdot x|}$$

where $H \cdot x = \{h \cdot x \mid h \in H\}$.

(Note that since H is symmetric, we have $1 \in K \cap H^{(2)}$.)

Proof. As in the classical proof of the orbit-stabilizer theorem, we consider the orbit map, but restricted to H

$$\phi : \begin{cases} H & \longrightarrow & X \\ h & \mapsto & h \cdot x \end{cases}$$

and we use it to count the number of elements in H: we have

$$|H| = \sum_{y \in \phi(H)} |\phi^{-1}(y)|,$$

and the point is that $\phi(H) = H \cdot x$ on the one hand, and

$$|\phi^{-1}(y)| \leqslant |K \cap H^{(2)}|$$

for all y, since if $y = \phi(h_0)$, with $h_0 \in H$, all elements $h \in H$ with $\phi(h) = y$ satisfy $hh_0^{-1} \in K \cap H^{(2)}$. Hence we get

$$|H| \leqslant |H \cdot x||K \cap H^{(2)}|,$$

as claimed. □

Finally, a last lemma shows that if a subset H has small tripling constant "in a subgroup", then H itself has small tripling (in the language of approximate groups, it is a special case of the fact that the intersection of two approximate groups is still one).

Lemma 3.4. *Let G be a finite group, $K \subset G$ a subgroup, and $H \subset G$ an arbitrary symmetric subset. We have*

$$\frac{|H^{(4)}|}{|H|} \geqslant \frac{|H^{(3)} \cap K|}{|H^{(2)} \cap K|}.$$

Proof. Let $X \subset G/K$ be the set of cosets of K intersecting H:

$$X = \{xK \in G/K \mid xK \cap H \neq \emptyset\}.$$

Let $xK \in X$ be given, and $xk = h \in xK \cap H$. Then all the elements xkg are distinct for $g \in K$, and they are in $xK \cap H^{(4)}$ if $g \in K \cap H^{(3)}$, so that

$$|xK \cap H^{(4)}| \geqslant |K \cap H^{(3)}|$$

for any $xK \in X$, and (cosets being disjoint)

$$|H^{(4)}| \geqslant |X||K \cap H^{(3)}|.$$

Similarly, for any $xK \in X$, fixing $g_0 \in xK \cap H$, we find that $g_0 g^{-1} \in K \cap H^{(2)}$ for any $g \in xK \cap H$, hence

$$|xK \cap H| \leqslant |K \cap H^{(2)}|.$$

Since

$$|H| = \sum_{xK \in X} |H \cap xK| \leqslant |X||K \cap H^{(2)}|,$$

we derive

$$|H^{(4)}| \geqslant \frac{|H|}{|K \cap H^{(2)}|}|K \cap H^{(3)}|.$$

□

We will use classical structural definitions and facts about finite groups of Lie type. In particular, a regular semisimple element $g \in \mathbf{G} = \mathrm{SL}_2(\bar{\mathbf{F}}_p)$ is a semisimple element with distinct eigenvalues. The centralizer of such an element is a maximal torus in $\mathbf{G}$. For any subset $H \subset \mathbf{G}$, we write H_{reg} for the set of the regular semisimple elements in H. A maximal torus $T \subset G = \mathrm{SL}_2(\mathbf{F}_p)$ is the intersection $G \cap \mathbf{T}$, where $\mathbf{T}$ is a maximal torus of $\mathbf{G}$ which is stable under the Frobenius automorphism σ. Here are the basic properties of regular semisimple elements and their centralizers; these are all standard facts, and we omit the proofs.

Proposition 3.5. *Fix a prime number p and let $G = \mathrm{SL}_2(\mathbf{F}_p)$, $\mathbf{G} = \mathrm{SL}_2(\bar{\mathbf{F}}_p)$.*

(1) *A regular semisimple element $x \in \mathbf{G}$ is contained in a* unique *maximal torus $\mathbf{T}$, namely its centralizer $\mathbf{T} = C_{\mathbf{G}}(x)$. In particular, if $\mathbf{T}_1 \neq \mathbf{T}_2$ are two maximal tori, we have*

$$\mathbf{T}_{1,reg} \cap \mathbf{T}_{2,reg} = \emptyset. \tag{3.2}$$

(2) *If $\mathbf{T} \subset \mathbf{G}$ is a maximal torus, we have*

$$|\mathbf{T}_{nreg}| = |\mathbf{T} - \mathbf{T}_{reg}| = 2.$$

(3) *For any maximal torus $\mathbf{T}$, the normalizer $N_{\mathbf{G}}(\mathbf{T})$ contains $\mathbf{T}$ as a subgroup of index 2. Similarly, for any maximal torus $T \subset G$, $N_G(T)$ contains T as a subgroup of index 2, and in particular*

$$2(p-1) \leqslant |N_G(T)| \leqslant 2(p+1).$$

(4) *The conjugacy class* $\mathbf{Cl}(g)$ *of a regular semisimple element* $g \in \mathbf{G}$ *is the set of all* $x \in \mathbf{G}$ *such that* $\mathrm{Tr}(x) = \mathrm{Tr}(g)$. *The set of elements in* $\mathbf{G}$ *which are not regular semisimple is the set of all* $x \in \mathbf{G}$ *such that* $\mathrm{Tr}(x)^2 = 4$.

Finally, (a variant of) the following concept was introduced under different names and guises by Helfgott, Pyber-Szabó, and Breuillard-Green-Tao. We chose the name from the last team.

Definition 3.6 (A set involved with a torus)**.** Let p be a prime number, $H \subset \mathrm{SL}_2(\mathbf{F}_p)$ a finite set and $\mathbf{T} \subset \mathrm{SL}_2(\bar{\mathbf{F}}_p)$ a maximal torus. Then H *is involved with* $\mathbf{T}$, or $\mathbf{T}$ with H, if and only if $\mathbf{T}$ is σ-invariant and H contains a regular semisimple element of $\mathbf{T}$ with non-zero trace, i.e., $H \cap \mathbf{T}_{sreg} \neq \emptyset$ where the superscript "sreg" restricts to regular semisimple elements with non-zero trace.

Remark 3.7. The twist in this definition, compared with the one in [16] or [3], is that we insist on having non-zero trace. This will be helpful later on, as it will eliminate a whole subcase in the key estimate (the proof of Proposition 3.11), and lead to a shorter proof, with better explicit constants. However, this restriction is not really essential in the greater scheme of things, and it would probably not be a good idea to do something similar for more general groups.

The alternative $H^{(3)} = \mathrm{SL}_2(\mathbf{F}_p)$ in Helfgott's growth theorem will be obtained as a corollary of the Gowers-Nikolov-Pyber "quasi-random groups" argument (see [6] and [14]).

Proposition 3.8. *For a prime* $p \geqslant 3$, *if a subset* $H \subset \mathrm{SL}_2(\mathbf{F}_p)$ *satisfies*

$$|H| \geqslant 2|\,\mathrm{SL}_2(\mathbf{F}_p)|^{8/9},$$

we have $H^{(3)} = \mathrm{SL}_2(\mathbf{F}_p)$.

For a proof, see, e.g., [11, §4.5].

3.2. **Escape from subvarieties and non-concentration lemmas.** Two important tools in the proof of growth theorems for linear groups are estimates for escape from subvarieties and for non-concentration in subvarieties. We state and prove in this section the special cases which we need for the explicit proof of Helfgott's Theorem. The reader may wish to look only at the statements and skip afterwards to the next section to see how they are used.

Lemma 3.9 (Escape)**.** *Let* $p \geqslant 7$ *be a prime number and let* $H \subset \mathrm{SL}_2(\mathbf{F}_p)$ *be a symmetric generating set with* $1 \in H$. *Then* $H^{(3)}_{sreg} \neq \emptyset$, *i.e., the three-fold product set* $H^{(3)}$ *contains a regular semisimple element with non-zero trace.*[7]

The general non-concentration inequalities are now often called "Larsen-Pink inequalities", since the first versions appeared in the work of Larsen and Pink [12] on finite subgroups of linear groups. "Approximate" versions occur in the work of Hrushovski [8] and Breuillard-Green-Tao [3], with closely related results found in that of Pyber and Szabó [16].

Theorem 3.10 (Non-concentration inequality)**.** *Let* $p \geqslant 3$ *be a prime number and let* $g \in \mathrm{SL}_2(\mathbf{F}_p) = G$ *be a regular semisimple element with non-zero trace. Let* $\mathbf{Cl}(g) \subset \mathrm{SL}_2(\bar{\mathbf{F}}_p) = \mathbf{G}$ *be the conjugacy class of* g. *If* $H \subset G$ *is a symmetric generating set containing* 1, *we have*

$$|\,\mathbf{Cl}(g) \cap H| \leqslant 7\alpha^{2/3}|H|^{2/3} \tag{3.3}$$

[7] The condition $p \geqslant 7$ is sharp, see [11, Example 4.6.13] for an example.

where $\alpha = \mathrm{trp}(H)$ *is the tripling constant of* H*,* unless

$$\alpha > |H|^{1/28}. \tag{3.4}$$

From this last fact, we will deduce the following dichotomy, which is the precise tool used in the next section to prove Helfgott's Theorem.

Proposition 3.11 (Involving dichotomy)**.** (1) *For all prime number* p*, all subsets* $H \subset \mathrm{SL}_2(\mathbf{F}_p)$ *and all maximal tori* $\mathbf{T} \subset \mathrm{SL}_2(\bar{\mathbf{F}}_p)$*, if* $\mathbf{T}$ *and* H *are not involved, we have*

$$|H \cap \mathbf{T}| \leqslant 4.$$

(2) *If* $p \geqslant 3$ *and* $H \subset \mathrm{SL}_2(\mathbf{F}_p) = G$ *is a symmetric generating set containing* 1*, we have*

$$|\mathbf{T}_{reg} \cap H^{(2)}| \geqslant 14^{-1}\alpha^{-4}|H|^{1/3} \tag{3.5}$$

for any maximal torus $\mathbf{T} \subset \mathrm{SL}_2(\bar{\mathbf{F}}_p)$ *which is involved with* $\mathbf{T}$*, where* $\alpha = \mathrm{trp}(H)$*, unless*

$$\alpha \geqslant |H|^{1/168}. \tag{3.6}$$

Proof. (1) is obvious, since $|\mathbf{T} - \mathbf{T}_{reg}| \leqslant 2$ and there are also at most two elements of trace 0 in $\mathbf{T}$ (as one can check quickly).

For (2), we apply the orbit-stabilizer theorem. Let $T = \mathbf{T} \cap \mathbf{G}$ be a maximal torus in G. Fixing any $g \in T_{reg}$, we have $T = C_G(g)$, the stabilizer of g in G for its conjugacy action on itself. We find that

$$|\mathbf{T} \cap H^{(2)}| \geqslant \frac{|H|}{|\{hgh^{-1} \mid h \in H\}|} \tag{3.7}$$

for any symmetric subset H. Since H is involved with $\mathbf{T}$, we can select g in $T_{sreg} \cap H$ in this inequality, and the denominator on the right becomes

$$|\{hgh^{-1} \mid h \in H\}| \leqslant |H^{(3)} \cap \mathrm{Cl}(g)| \leqslant |H^{(3)} \cap \mathbf{Cl}(g)|$$

where $\mathrm{Cl}(g)$ is the conjugacy class of g in G. Applying the Larsen-Pink inequality to $H^{(3)}$, with tripling constant bounded by α^6 (by Ruzsa's Lemma), we obtain the lower bound

$$|\mathbf{T} \cap H^{(2)}| \geqslant \frac{|H|}{|H^{(3)} \cap \mathbf{Cl}(g)|} \geqslant 7^{-1}\alpha^{-4}|H|^{1/3},$$

unless $\alpha = \mathrm{trp}(H) \geqslant |H|^{1/168}$. In the first case, we get

$$|\mathbf{T}_{reg} \cap H^{(2)}| \geqslant 14^{-1}\alpha^{-4}|H|^{1/3},$$

unless

$$7^{-1}\alpha^{-4}|H|^{1/3} \leqslant 2$$

since there are only two elements of $\mathbf{T} \cap H^{(2)}$ which are not regular. This last alternative gives

$$\alpha \geqslant \tfrac{1}{2}|H|^{1/12}$$

which we see is a stronger conclusion than (3.6) (precisely, it is strictly stronger if $|H| > 2^{13}$, but in the other case the lower bound $\mathrm{trp}(H) \geqslant \sqrt{2}$ from Lemma 3.2 is already a better result.) Hence Proposition 3.11 is proved. ∎

Now we prove the escape and non-concentration results.

Proof of Lemma 3.9. The basic point that allows us to give a quick proof is that the set $\mathbf{N} = \mathbf{G} - \mathbf{G}_{reg}$ of elements which are not regular semisimple is invariant under conjugation, and is the set of all matrices with trace equal to 2 or -2. It is precisely the union of the two central elements ± 1 and the four conjugacy classes of

$$u = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad v = \begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}, \quad u' = \begin{pmatrix} 1 & \varepsilon \\ 0 & 1 \end{pmatrix}, \quad v = \begin{pmatrix} -1 & \varepsilon \\ 0 & -1 \end{pmatrix}$$

(where $\varepsilon \in \mathbf{F}_p^\times$ is a fixed non-square) while elements of trace 0 are the conjugates of

$$g_0 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

(these are all standard facts.)

We next note that, if the statement of the lemma fails for a given H, it also fails for every conjugates of H, and that this allows us to normalize at least one element to a specific representative of its conjugacy class. It is convenient to argue by contradiction, though this is somewhat cosmetic. So we assume that $H_{nreg}^{(3)}$ is empty and $p \geqslant 7$, and will derive a contradiction.

We distinguish two cases. In the first case, we assume that H contains one element of trace ± 2 which is not ± 1. The observation above shows that we can assume that one of u, v, u', v' is in H, and we deal with the case $u \in H$ (the others being exactly analogue.)

Since H is a symmetric generating set, it must contain some element

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

with $c \neq 0$, since otherwise, all elements of H would be upper-triangular, and H would not generate $\mathrm{SL}_2(\mathbf{F}_p)$. Then $H^{(3)}$ contains ug, u^2g, $u^{-1}g$, $u^{-2}g$, which have traces, respectively, equal to $\mathrm{Tr}(g) + c$, $\mathrm{Tr}(g) + 2c$, $\mathrm{Tr}(g) - c$, $\mathrm{Tr}(g) - 2c$. Since $c \neq 0$, and p is not 2 or 3, we see that these traces are distinct, and since there are 4 of them, one at least is not in $\{-2, 0, 2\}$, which contradicts our assumption.

In the second case, all elements of H except ± 1 have trace 0. We split in two subcases, but depending on properties of $\mathbf{F}_p$.

The first one is when -1 is *not* a square in $\mathbf{F}_p$. Conjugating again, we can assume that $g_0 \in H$. Because H generates $\mathrm{SL}_2(\mathbf{F}_p)$, we claim that there must exist a matrix

$$g = \begin{pmatrix} a & b \\ c & -a \end{pmatrix}$$

in H with (i) $a \neq 0$; (ii) $b \neq c$. Indeed if all elements $\neq \pm 1$ of H are of the form

$$g = \begin{pmatrix} 0 & -c^{-1} \\ c & 0 \end{pmatrix},$$

we can find such an element with $c \neq \pm 1$ (i.e., $g \neq \pm g_0$), since otherwise H is not a generating set; then the trace of $g_0 g$ is $c + c^{-1}$, which is not in $\{-2, 0, 2\}$ (non-zero because -1 is not a square in our first subcase), so $H_{nreg}^{(2)} \neq \emptyset$, which we excluded. So all elements of H, except for ± 1 and $\pm g_0$ are of the type

$$g = \begin{pmatrix} a & b \\ c & -a \end{pmatrix},$$

with $a \neq 0$. Then, if all these satisfied $b = c$, it would follow that H is contained in the normalizer of a non-split maximal torus, again contradicting the assumption that H is a generating set.

Now we argue with g as above. We have

$$g_0 g = \begin{pmatrix} c & -a \\ -a & -b \end{pmatrix} \in H^{(2)},$$

with non-zero trace $t = c-b$. Moreover, if $t = 2$, i.e., $c = b+2$, the condition $\det(g_0 g) = 1$ implies

$$-2b - b^2 - a^2 = 1$$

or $(b+1)^2 = -a^2$. Similarly, if $t = -2$, we get $(b-1)^2 = -a^2$. Since $a \neq 0$, it follows in both cases that -1 is a square in $\mathbf{F}_p$, which contradicts our assumption in the first subcase.

Now we come to the second subcase when $-1 = z^2$ is a square in $\mathbf{F}_p$. We can then diagonalize g_0 over $\mathbf{F}_p$, and conjugating again, this means we can assume that H contains

$$g_0' = \begin{pmatrix} z & 0 \\ 0 & -z \end{pmatrix}$$

as well as some other matrix

$$g' = \begin{pmatrix} a & b \\ c & -a \end{pmatrix}$$

(the values of a, b, c are not the same as before; we are still in the case when every element of H has trace 0 except for ± 1).

Now the trace of $g_0' g' \in H^{(2)}$ is $2za$. But we can find g' with $a \neq 0$, since otherwise H would again not be a generating set, being contained in the normalizer of the diagonal (split) maximal torus, and so this trace is non-zero.

The condition $2za = \pm 2$ would give $za = \pm 1$, which leads to $-a^2 = 1$. But since $1 = \det(g') = -a^2 - bc$, we then get $bc = 0$ for all elements of H. Finally, if all elements of H satisfy $b = 0$, the set H would be contained in the subgroup of upper-triangular matrices. So we can find a matrix in H with $b \neq 0$, hence $c = 0$. Similarly, we can find another

$$g'' = \begin{pmatrix} a & 0 \\ c & -a \end{pmatrix}$$

in H with $c \neq 0$. Taking into account that $z^2 = -1$, computing the traces of $g'g''$ and of $g_0 g' g''$ gives

$$bc - 2, \qquad bcz$$

respectively. If $bc = 2$, the third trace (of an element in $H^{(3)}$) is $2z \notin \{0, 2, -2\}$ since $p \neq 2$, and if $bc = 4$, it is $4z \notin \{0, 2, -2\}$ since $p \neq 5$. And of course if $bc \notin \{2, 4\}$, the first trace is already not in $\{-2, 0, 2\}$. So we are done... □

For the proof of Theorem 3.10, we will use the method suggested by Larsen and Pink at the beginning of [12, §4]. We consider the map

$$\phi \begin{cases} \mathbf{Cl}(g) \times \mathbf{Cl}(g) \times \mathbf{Cl}(g) & \longrightarrow \quad \mathbf{G} \times \mathbf{G} \\ \qquad (x_1, x_2, x_3) & \mapsto \quad (x_1 x_2, x_1 x_3) \end{cases}$$

and we note that for $(x_1, x_2, x_3) \in (\mathbf{Cl}(g) \cap H)^3$, we have $\phi(x_1, x_2, x_3) \in H^{(2)}$. We then hope that the fibers $\phi^{-1}(y_1, y_2)$ of ϕ are all finite with size bounded independently of $(y_1, y_2) \in \mathbf{G} \times \mathbf{G}$, say of size at most $c_1 \geqslant 1$. The reason behind this hope is that $\mathbf{Cl}(g)^3$ and $\mathbf{G}^2$ have the same dimension, and hence unless something special happens, we would expect the fibers to have dimension 0, which corresponds to having fibers of bounded size since everything is defined using polynomial equations.

If this hope turns out to be justified, we can count $|\mathbf{Cl}(g)\cap H|$ by summing according to the values of ϕ: denoting $Z = (\mathbf{Cl}(g)\cap H)^3$ and $W = \phi(Z) = \phi((\mathbf{Cl}(g)\cap H)^3)$, we have

$$|\mathbf{Cl}(g)\cap H|^3 = |Z| = \sum_{(y_1,y_2)\in W} |\phi^{-1}(y_1,y_2)\cap Z|$$

which – under our optimistic assumption – leads to the estimate

$$|\mathbf{Cl}(g)\cap H|^3 \leqslant c_1|W| \leqslant c_1|H^{(2)}|^2 \leqslant c_1\alpha^2|H|,$$

which has the form we want.

To implement this – and solve the complications that arise –, we are led to analyze the fibers of the map ϕ. The resulting computations were explained to the author by R. Pink, and start with an easy observation:

Lemma 3.12. *Let k be any field, and let $G = \mathrm{SL}_2(k)$. Let $C \subset G$ be a conjugacy class, and define*

$$\phi\begin{cases} C^3 & \longrightarrow & G^2 \\ (x_1,x_2,x_3) & \mapsto & (x_1x_2, x_1x_3) \end{cases}.$$

Then for any $(y_1,y_2)\in G\times G$, we have a bijection

$$\begin{cases} C\cap y_1C^{-1}\cap y_2C^{-1} & \longrightarrow & \phi^{-1}(y_1,y_2) \\ x_1 & \mapsto & (x_1, x_1^{-1}y_1, x_1^{-1}y_2) \end{cases}.$$

In particular, if $k = \bar{\mathbf{F}}_p$ and C is a regular semisimple conjugacy class, we have a bijection

$$\phi^{-1}(y_1,y_2) \longrightarrow C\cap y_1C\cap y_2C.$$

Proof. Taking x_1 as a parameter, any (x_1,x_2,x_3) with $\phi(x_1,x_2,x_3) = (y_1,y_2)$ can certainly be written $(x_1, x_1^{-1}y_1, x_1^{-1}y_2)$. Conversely, such an element in $\mathrm{SL}_2(k)^3$ really belongs to C^3 (hence to the fiber) if and only if $x_1\in C$, $x_1^{-1}y_1\in C$, $x_1^{-1}y_2\in C$, i.e., if and only if $x_1 \in C\cap y_1C^{-1}\cap y_2C^{-1}$, which proves the first part.

For the second part, we need only notice that if C is a regular semisimple conjugacy class, say that of g, then $C = C^{-1}$ because g^{-1} has the same characteristic polynomial as g, hence is conjugate to g. ☐

We are now led to determine when an intersection of the form $C\cap y_1C\cap y_2C$ can be infinite. The answer is as follows, and it is one place where the use of the infinite group $\mathrm{SL}_2(\bar{\mathbf{F}}_p)$ is significant:

Lemma 3.13 (Pink). *Let k be an algebraically closed field of characteristic not equal to 2, and let $g\in \mathrm{SL}_2(k)$ be a regular semisimple element, C the conjugacy class of g. For y_1, $y_2\in G$, the intersection $X = C\cap y_1C\cap y_2C$ is finite, containing at most two elements, unless one of the following cases holds:*

(1) *We have $y_1 = 1$, or $y_2 = 1$ or $y_1 = y_2$.*

(2) *There exists a conjugate $\mathbf{B} = x\mathbf{B}_0x^{-1}$ of the subgroup*

$$\mathbf{B}_0 = \left\{\begin{pmatrix} a & b \\ 0 & a^{-1}\end{pmatrix}\right\} \subset \mathrm{SL}_2(k)$$

and an element $t\in\mathbf{B}\cap C$ such that

$$y_1, y_2 \in \mathbf{U}\cup t^2\mathbf{U} \tag{3.8}$$

where

$$\mathbf{U} = x\mathbf{U}_0x^{-1}, \qquad \mathbf{U}_0 = \left\{\begin{pmatrix} 1 & b \\ 0 & 1\end{pmatrix}\right\} \subset \mathbf{B}_0.$$

In that case, we have $X \subset C \cap \mathbf{B}$.
(3) *The trace of g is* 0.

The proof will be given at the end of this section: it is mostly computational. Before coming back to the proof of Theorem 3.10, we state and prove another preliminary lemma, which is another case of non-concentration inequalities.

Lemma 3.14. *For a prime p and $\gamma \in \mathbf{F}_p^\times$, define*

$$C_\gamma = \Bigl\{ \begin{pmatrix} \gamma & t \\ 0 & \gamma^{-1} \end{pmatrix} \mid t \in \bar{\mathbf{F}}_p \Bigr\}.$$

For any $p \geqslant 3$, any $\gamma \in \mathbf{F}_p^\times$, any $x \in \mathrm{SL}_2(\bar{\mathbf{F}}_p)$ and any symmetric generating set H of $\mathrm{SL}_2(\mathbf{F}_p)$ containing 1, we have

$$|H \cap xC_\gamma x^{-1}| = \Bigl| H \cap x\Bigl\{ \begin{pmatrix} \gamma & t \\ 0 & \gamma^{-1} \end{pmatrix} \mid t \in \mathbf{F}_p \Bigr\} x^{-1} \Bigr| \leqslant 2\alpha^2 |H|^{1/3}$$

where $\alpha = \mathrm{trp}(H)$.

Proof. We first deal with the fact that x and γ are not necessarily in $\mathrm{SL}_2(\mathbf{F}_p)$. We have $xC_\gamma x^{-1} \cap \mathrm{SL}_2(\mathbf{F}_p) \subset x\mathbf{B}_0 x^{-1} \cap \mathrm{SL}_2(\mathbf{F}_p)$, and there are three possibilities for the latter: either $x\mathbf{B}_0 x^{-1} \cap \mathrm{SL}_2(\mathbf{F}_p) = 1$, or $x\mathbf{B}_0 x^{-1} \cap \mathrm{SL}_2(\mathbf{F}_p) = T$ is a non-split maximal torus of $\mathrm{SL}_2(\mathbf{F}_p)$, or $x\mathbf{B}_0 x^{-1} \cap \mathrm{SL}_2(\mathbf{F}_p) = B$ is an $\mathrm{SL}_2(\mathbf{F}_p)$-conjugate of the group $B_0 = \mathbf{B}_0 \cap \mathrm{SL}_2(\mathbf{F}_p)$ of upper-triangular matrices (this is once more a standard property of linear algebraic groups over finite fields). In this last case, we can assume that $x \in \mathrm{SL}_2(\mathbf{F}_p)$ and $\gamma \in \mathbf{F}_p$. In the first, of course, there is nothing to do. And as for the second, note that γ and γ^{-1} are the eigenvalues of any element in $\mathrm{SL}_2(\mathbf{F}_p) \cap xC_\gamma x^{-1}$, and there are at most two elements in a maximal torus with given eigenvalues. A fortiori, we have $|H \cap xC_\gamma x^{-1}| \leqslant 2 \leqslant 2\alpha^2 |H|^{1/3}$ in that case.

Thus we are left with the situation where $x \in \mathrm{SL}_2(\mathbf{F}_p)$. Using $\mathrm{SL}_2(\mathbf{F}_p)$-conjugation, it is enough to deal with the case $x = 1$. Then either the intersection is empty (and the result true) or we can fix

$$g_0 = \begin{pmatrix} \gamma & t_0 \\ 0 & \gamma^{-1} \end{pmatrix} \in H \cap C_\gamma,$$

and observe that for any $g \in H \cap C_\gamma$, we have

$$g_0^{-1} g \in H^{(2)} \cap C_1,$$

hence

$$|H \cap C_\gamma| \leqslant |H^{(2)} \cap C_1| = |H^{(2)} \cap \mathbf{U}_0|,$$

which reduces further to the case $\gamma = 1$.

In that case we have another case of the Larsen-Pink non-concentration inequality, in that case in a one-dimensional variety. There is here also a rather short proof: we fix any element $h \in H$ such that h is *not* in $\mathbf{B}_0$, i.e.

$$h = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

with $c \neq 0$. This element exists, because otherwise $H \subset \mathbf{B} \cap \mathrm{SL}_2(\mathbf{F}_p)$ would not be a generating set of $\mathrm{SL}_2(\mathbf{F}_p)$.

Now consider the multiplication map

$$\psi : \begin{cases} \mathbf{U}^* \times \mathbf{U}^* \times \mathbf{U}^* & \longrightarrow \quad \mathbf{G} \\ \quad (u_1, u_2, u_3) & \mapsto \quad u_1 h u_2 h^{-1} u_3 \end{cases}$$

where $\mathbf{U}^* = \mathbf{U}_0 - 1$ (we explain below why we do not use $\mathbf{U}_0^3$ as domain).

Note that since $h \in H$, we have $\psi((\mathbf{U}^* \cap H^{(2)})^3) \subset H^{(8)}$. Crucially, we claim that for *any* $x \in \mathbf{G}$, the fiber $\psi^{-1}(x)$ is either empty or reduced to a single element! If this is true, we get as before

$$|\mathbf{U}^* \cap H^{(2)}|^3 \leqslant |H^{(8)}| \leqslant \alpha^6 |H|,$$

and therefore

$$|\mathbf{U}_0 \cap H^{(2)}| = |\mathbf{U}^* \cap H^{(2)}| + 1 \leqslant 2\alpha^2 |H|^{1/3},$$

which is the result.

To check the claim, we compute. Precisely, if

$$u_i = \begin{pmatrix} 1 & t_i \\ 0 & 1 \end{pmatrix} \in \mathbf{U}^*,$$

a matrix multiplication leads to

$$\psi(u_1, u_2, u_3) = \begin{pmatrix} -t_1 t_2 c^2 - t_2 ad & \star \\ -t_2 c^2 & \star \end{pmatrix},$$

and in order for this to be a fixed matrix x, we see that t_2 (i.e., u_2) is uniquely determined (since $c \neq 0$). Since u_2 is in $\mathbf{U}^*$, it is not 1, and this means that $t_2 \neq 0$ (ensuring this is the reason that ψ is defined using $\mathbf{U}^*$ instead of $\mathbf{U}_0$). Thus t_1 (i.e. u_1) is also uniquely determined, and finally

$$u_3 = (u_1 h u_2 h^{-1})^{-1} x$$

is uniquely determined. □

Proof of Theorem 3.10. We have g regular semisimple with $\mathrm{Tr}(g) \neq 0$. We define as above the map ϕ and denote

$$Z = (\mathbf{Cl}(g) \cap H)^3, \quad W = \phi(Z) = \phi((\mathbf{Cl}(g) \cap H)^3),$$

so that

$$|\,\mathbf{Cl}(g) \cap H|^3 = \sum_{(y_1, y_2) \in W} |\phi^{-1}(y_1, y_2) \cap Z| = S_0 + S_1 + S_2, \tag{3.9}$$

where S_i denotes the sum restricted to a subset $W_i \subset W$, W_0 being the subset where the fiber has order at most 2, while W_1, W_2 correspond to those (y_1, y_2) where cases (1) and (2) of Lemma 3.13 hold. Precisely, we do not put into W_2 the (y_1, y_2) for which both cases (1) and (2) are valid, e.g., $y_1 = 1$, and we *add* to W_1 the cases where $y_1 = -1$, which may otherwise appear in Case (2). We will prove:

$$S_0 \leqslant 2|H^{(2)}|^2 \leqslant 2\alpha^2 |H|^2, \qquad S_1 \leqslant 4|H^{(3)}|^2 \leqslant 4\alpha^2 |H|^2,$$
$$S_2 \leqslant 16\alpha^{34/3} |H|^{5/3}.$$

Assuming this, we get immediately

$$|\,\mathbf{Cl}(g) \cap H| \leqslant 6^{2/3} \alpha^{2/3} |H|^{2/3} + 2^{4/3} \alpha^{34/9} |H|^{5/9}$$

from (3.9). Now either the second term is smaller than the first, and we get (3.3) (since $2 \cdot 6^{2/3} < 7$), or

$$2^{4/3} \alpha^{34/9} |H|^{5/9} > 6^{2/3} \alpha^{2/3} |H|^{2/3} > 2^{4/3} \alpha^{2/3} |H|^{2/3},$$

which gives

$$\alpha > |H|^{1/28},$$

the second alternative (3.4) of Theorem 3.10, which is therefore proved.

We now check the bounds on S_i. The case of S_0 follows by the fact that the fibers over W_0 have at most two elements, hence also their intersection with Z, and that $|W_0| \leqslant |W| \leqslant |H^{(2)}|^2$.

The case of S_1 splits into four almost identical subcases, corresponding to $y_1 = 1$, $y_1 = -1$ (remember that we added this, borrowing it from Case (2)), $y_2 = 1$ or $y_1 = y_2$. We deal only with the first, say $S_{1,1}$: we have

$$S_{1,1} \leqslant \sum_{y_2 \in H^{(2)}} |\phi^{-1}(1, y_2) \cap Z|.$$

But using Lemma 3.12, we have

$$|\phi^{-1}(1, y_2) \cap Z| = |\{(x_1, x_1^{-1}, x_1^{-1} y_2) \in (\mathbf{Cl}(g) \cap H)^3\}| \leqslant |H^{(3)}|$$

for any given $y_2 \in H^{(2)}$, since x_1 determines the triple $(x_1, x_1^{-1}, x_1^{-1} y_2)$ and $x_1^{-1} = x_1^{-1} y_2 y_2^{-1} \in H^{(3)}$ for any such triple if $y_2 \in H^{(2)}$. Therefore

$$S_{1,1} \leqslant |H^{(2)}||H^{(3)}| \leqslant |H^{(3)}|^2,$$

and similarly for the other three cases.

Now for S_2. Here also we sum over y_1 first, which is $\neq \pm 1$ (by our definition of W_2). The crucial point is then that an element $y_1 \neq \pm 1$ is included in at most two conjugates of $\mathbf{B}_0$. Hence, up to a factor 2, the choice of y_1 fixes that of the relevant conjugate $\mathbf{B}$ for which Case (2) applies. Next we observe that $C_{\mathbf{B}} = \mathbf{Cl}(g) \cap \mathbf{B}$ is a conjugate of the union

$$C_\alpha \cup C_{\alpha^{-1}},$$

where, as in Lemma 3.14, we define

$$C_\alpha = \left\{ \begin{pmatrix} \alpha & t \\ 0 & \alpha^{-1} \end{pmatrix} \mid t \in \bar{\mathbf{F}}_p \right\},$$

and α is such that $\alpha + \alpha^{-1} = \mathrm{Tr}(g)$. Given $y_1 \in H^{(2)}$ and $\mathbf{B}$ containing y_1, we have by (3.8)

$$y_2 \in (H^{(2)} \cap \mathbf{U}) \cup (H^{(2)} \cap t^2 \mathbf{U})$$

for some $t \in C_{\mathbf{B}}$. We note that $t^2\mathbf{U}$ is itself conjugate to C_{α^2} or $C_{\alpha^{-2}}$.

Then the size of the fiber $\phi^{-1}(y_1, y_2) \cap Z$ is determined by the number of possibilities for x_1. As the latter satisfies

$$x_1 \in C_{\mathbf{B}} \cap H,$$

we see that we must estimate the size of intersections of the type

$$H \cap C_\gamma, \;\; H^{(2)} \cap C_\gamma$$

for some fixed $\gamma \in \mathbf{F}_p^\times$, as this will lead us to estimates for the number of possibilities for y_2 as well as x_1. Using twice Lemma 3.14, we get

$$|\{y_2 \mid (y_1, y_2) \in W_2\}| \leqslant 8\mathrm{trp}(H^{(2)})^2 |H^{(2)}|^{1/3} \leqslant 8\alpha^{25/3}|H|^{1/3},$$

(the factor 8 accounts for the two possible choices of $\mathbf{B}$ and the two "components" for y_2, and the factor 2 in the lemma) and

$$|\phi^{-1}(y_1, y_2) \cap Z| \leqslant 4\alpha^2 |H|^{1/3}.$$

This gives

$$S_2 \leqslant 16\alpha^{31/3}|H|^{2/3}|H^{(2)}| \leqslant 16\alpha^{34/3}|H|^{5/3},$$

as claimed. □

There now only remains to prove Lemma 3.13.

Proof of Lemma 3.13. It will be convenient to compute the intersection $C\cap y_1^{-1}C\cap y_2^{-1}C$ instead of $C\cap y_1C\cap y_2C$, a change of notation whichs is innocuous.

The computation is then based on a list of simple checks. We can assume that the regular semisimple element g is

$$g=\begin{pmatrix}\alpha & 0\\ 0 & \alpha^{-1}\end{pmatrix}$$

where $\alpha^4\neq 1$, because $\alpha=\pm 1$ implies that g is not regular semisimple, and α a fourth root of unity implies that $\mathrm{Tr}(g)=0$, which is the third case of the lemma (recall that k is assumed to be algebraically closed). Thus the conjugacy class C is the set of matrices of trace equal to $t=\alpha+\alpha^{-1}$.

The only trick involved is that, for any $y_1\in\mathrm{SL}_2(k)$ and $x\in\mathrm{SL}_2(k)$, we have

$$C\cap(xy_1x^{-1})^{-1}C=x(x^{-1}C\cap y_1^{-1}x^{-1}C)=x(C\cap y_1^{-1}C)x^{-1}$$

since $x^{-1}C=Cx^{-1}$, by definition of conjugacy classes. This means we can compute $C\cap y_1^{-1}C$, up to conjugation, by looking at $C\cap(y_1')^{-1}C$ for any y_1' in the conjugacy class of y_1. In particular, of course, determining whether $C\cap y_1^{-1}C$ is infinite or not only depends on the conjugacy class of y_1.

The conjugacy classes in $\mathrm{SL}_2(k)$ are well-known. We will run through representatives of these classes in order, and determine the corresponding intersection $C\cap y_1^{-1}C$. Then, to compute $C\cap y_1^{-1}C\cap y_2^{-1}C$, we take an element x in $C\cap y_1^{-1}C$, compute y_2x, and $C\cap y_1^{-1}C\cap y_2^{-1}C$ corresponds to those x for which the trace of y_2x is also equal to t.

We assume $y_1\neq\pm 1$. Then we distinguish four cases:

$$y_1=\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix},\quad y_1=\begin{pmatrix}-1 & 1\\ 0 & -1\end{pmatrix},$$
$$y_1=\begin{pmatrix}\beta & 0\\ 0 & \beta^{-1}\end{pmatrix},\quad \beta\neq\pm 1,\beta\neq\alpha^{\pm 2} \tag{3.10}$$
$$y_1=\begin{pmatrix}\alpha^2 & 0\\ 0 & \alpha^{-2}\end{pmatrix}.$$

We claim that $D=C\cap y_1^{-1}C$ is then given, respectively, by the sets containing all matrices of the following forms, parameterized by an element $a\in k$ (with $a\neq 0$ in the third case):

$$\begin{pmatrix}\alpha & a\\ 0 & \alpha^{-1}\end{pmatrix}\text{ or }\begin{pmatrix}\alpha^{-1} & a\\ 0 & \alpha\end{pmatrix}, \tag{3.11}$$
$$\begin{pmatrix}a & (-a^2+at-1)/(2t)\\ 2t & t-a\end{pmatrix},$$

$$\frac{1}{\beta+1}\begin{pmatrix}t & (\beta-\alpha^2)a\\ -(\beta-\alpha^{-2})a^{-1} & t\beta\end{pmatrix}, \tag{3.12}$$

$$\begin{pmatrix}\alpha^{-1} & a\\ 0 & \alpha\end{pmatrix}\text{ or }\begin{pmatrix}\alpha^{-1} & 0\\ a & \alpha\end{pmatrix}. \tag{3.13}$$

Let us check, for instance, the third and fourth cases (cases (1) and (2) are left as exercise), which we can do simultaneously, taking y_1 as in (3.10) but without assuming $\beta\neq\alpha^{\pm 2}$. For

$$x=\begin{pmatrix}a & b\\ c & d\end{pmatrix}\in C,$$

we compute

$$y_1 x = \begin{pmatrix} \beta a & \beta b \\ \beta^{-1} c & \beta^{-1} d. \end{pmatrix}$$

This matrix belongs to C if and only if $\beta a + \beta^{-1} d = t = a + d$. This means that (a, d) is a solution of the linear system

$$\begin{cases} a + d = t \\ \beta a + \beta^{-1} d = t, \end{cases}$$

of determinant $\beta^{-1} - \beta \neq 0$, so that we have

$$a = \frac{t}{\beta + 1}, \qquad d = \frac{\beta t}{\beta + 1}.$$

Write $c = c'/(\beta+1)$, $d = d'/(\beta+1)$; then the condition on c' and d' to have $\det(x) = 1$ can be expressed as

$$-c'd' = (\beta - \alpha^2)(\beta - \alpha^{-2}).$$

This means that either β is not one of α^2, α^{-2} (the third case), and then c and d are non-zero, and we can parametrize the solutions as in (3.12), or else (the fourth case) c or d must be zero, and then we get upper or lower-triangular matrices, as described in (3.13).

Now we intersect D (in the general case again) with $y_2^{-1} C$. We write

$$y_2 = \begin{pmatrix} x_1 & x_2 \\ x_3 & x_4 \end{pmatrix}.$$

We consider the first of our four possibilities now, so that $x \in D$ is upper-triangular with diagonal coefficients α, α^{-1} (as a set), see (3.11). We compute the trace of $y_2 x$, and find that is

$$a x_3 + x_1 \alpha + x_4 \alpha^{-1}, \text{ or } a x_3 + x_1 \alpha^{-1} + x_4 \alpha.$$

Thus, if $x_3 \neq 0$, there is at most one value of a for which the trace is t, i.e., $D \cap y_2^{-1} C$ has at most two elements (one for each form of the diagonal). If $x_3 = 0$, we find that x_1 is a solution of

$$\alpha x_1 + \alpha^{-1} x^{-1} = t,$$

or

$$\alpha x_1^{-1} + \alpha^{-1} x_1 = t,$$

for which the solutions are among 1, α^2 and α^{-2}, so that y_2 is upper-triangular with diagonal coefficients $(1, 1)$, (α^2, α^{-2}) or (α^{-2}, α^2), and this is one of the instances of Case (2) of Lemma 3.13.

Let us now consider the second of our four cases, leaving this time the third and fourth to the reader. Thus we take x as in (3.12), and compute the trace of $y_2 x$ as a function of a, which gives

$$\mathrm{Tr}(y_2 x) = -\frac{x_3}{2t} a^2 + \Big(x_1 - x_4 + \frac{x_3}{2}\Big) a + (x_4 + 2x_2) t.$$

The equation $\mathrm{Tr}(y_2 x) = t$ has therefore at most two solutions, unless $x_3 = 0$ and $x_4 = x_1$. In that case we have $x_4 = \pm 1$, and the constant term is equal to t if and only if $x_4 = 1$ and $x_2 = 0$ (so $y_2 = 1$) or $x_4 = \pm 1$ and $x_2 = 1$ (and then $y_2 = y_1$). Each of these possibilities corresponds to the exceptional situation of Case (1) of Lemma 3.13.

All in all, going through the remaining situations, we finish the proof. □

3.3. **Proof of Helfgott's Theorem.** We now prove Theorem 1.2. If $p \leqslant 5$, one checks numerically that trivial bounds already imply the theorem. So we assume that $p \geqslant 7$, which means that Lemma 3.9 is applicable. We will show that

$$\mathrm{trp}(H) \geqslant 2^{-1/2}|H|^{1/672} \tag{3.14}$$

for $p \geqslant 7$, unless $H^{(3)} = \mathrm{SL}_2(\mathbf{F}_p)$, where the latter case will arise by applying Proposition 3.8. Then using Lemma 3.2, we derive

$$\mathrm{trp}(H) \geqslant \max(2^{1/2}, 2^{-1/2}|H|^{1/672}) \geqslant |H|^{1/1344},$$

which is the precise form of Helfgott's Theorem we claimed.

We define $\tilde{H} = H^{(2)}$, so that (by Lemma 3.9) there exists at least one maximal torus $\mathbf{T}$ involved with $L = \tilde{H}^{(2)} = H^{(4)}$.

If, among all maximal tori involved with L, none satisfies (3.5), we obtain directly from Proposition 3.11 (applied to $\tilde{H}$ instead of H) the lower bound

$$\mathrm{trp}(\tilde{H}) \geqslant |\tilde{H}|^{1/168} \geqslant |H|^{1/168},$$

and since $\mathrm{trp}(\tilde{H}) \leqslant \alpha^4$ by Ruzsa's Lemma, we get

$$\alpha \geqslant |H|^{1/672} \geqslant 2^{-1/2}|H|^{1/672}, \tag{3.15}$$

which is (3.14).

Otherwise, we distinguish two cases.

Case (1). There exists a maximal torus $\mathbf{T}$ which is σ-invariant and is such that for any $g \in G$, the torus $g\mathbf{T}g^{-1}$ is involved with L.

As we can guess from (3.5) and (3.2), in that case, the set $\tilde{H}$ will tend to be rather large, so $|\tilde{H}|$ is close to $|G|$, *unless* the ratio $|L|/|\tilde{H}|$ is big, but then the tripling constant is even larger.

Precisely, writing $T = \mathbf{T} \cap G$, we note that the maximal tori

$$gTg^{-1} = (g\mathbf{T}g^{-1}) \cap G$$

are distinct for g taken among representatives of $G/N_G(T)$. Then we have the inequalities

$$|L^{(2)}| \geqslant \sum_{g \in G/N_G(T)} |L^{(2)} \cap g\mathbf{T}g^{-1}| \geqslant 7^{-1}\beta^{-4}|L|^{1/3}\frac{|G|}{|N_G(T)|} \geqslant 14^{-1}\beta^{-4}(p-1)^2|L|^{1/3}$$

where $\beta = \mathrm{trp}(L)$, since each $g\mathbf{T}g^{-1}$ is involved with L and distinct regular semisimple elements lie in distinct maximal tori.

Now we unwind this inequality in terms of H. We have $L^{(2)} = H^{(8)}$, so

$$|H| \geqslant \alpha^{-6}|L^{(2)}| \geqslant 14^{-1}\alpha^{-6}\beta^{-4}(p-1)^2|L|^{1/3} \geqslant 14^{-1}\alpha^{-6}\beta^{-4}(p-1)^2|H|^{1/3}$$

with $\alpha = \mathrm{trp}(H)$, by Ruzsa's Lemma. Furthermore, we have

$$\beta = \mathrm{trp}(L) = \mathrm{trp}(H^{(4)}) \leqslant \alpha^9$$

by Ruzsa's Lemma again, and hence the inequality gives the bound

$$|H| \geqslant 14^{-3/2}\alpha^{-63}(p-1)^3,$$

which for $p \geqslant 5$ implies $|H| \geqslant 100^{-1}\alpha^{-63}|G|$. But then either

$$\mathrm{trp}(H) = \alpha \geqslant 200^{-1/50}|G|^{1/567} \geqslant 2^{-1/2}|H|^{1/567}, \tag{3.16}$$

or else

$$|H| \geqslant 2|G|^{8/9},$$

which (via Proposition 3.8) are versions of the two alternatives we are seeking (in particular the first implies (3.14).)

Case (2). Since we know that *some* torus is involved with L, the complementary situation to Case (1) is that there exists a maximal torus $\mathbf{T}$ involved with $L = H^{(4)}$ and a conjugate $g\mathbf{T}g^{-1}$, for some $g \in G$, which is *not* involved. We are then going to get growth using Lemma 3.4. There is a first clever observation (the idea of which goes back to work of Glibichuk and Konyagin [5] on the "sum-product phenomenon"): one can assume, possibly after changing $\mathbf{T}$ and g, that g is in H.

Indeed, to check this claim, we start with $\mathbf{T}$ and h as above. Since H is a generating set, we can write

$$g = h_1 \cdots h_m$$

for some $m \geqslant 1$ and some elements $h_i \in H$. Now let $i \leqslant m$ be the smallest index such that the maximal torus

$$\mathbf{T}' = (h_{i+1} \cdots h_m)\mathbf{T}(h_{i+1} \cdots h_m)^{-1}$$

is involved with L. Taking $i = m$ means that $\mathbf{T}$ is involved with H, which is the case, and therefore the index i exists. Moreover $i \neq 0$, again by definition. It follows that

$$(h_i h_{i+1} \cdots h_m)\mathbf{T}(h_i h_{i+1} \cdots h_m)^{-1}$$

is not involved with L. But this means that we can replace $(\mathbf{T}, g)$ with $(\mathbf{T}', h_i)$, and since $h_i \in H$, this gives us the claim.

We now write h for the conjugator g such that L and the torus $\mathbf{S} = g\mathbf{T}g^{-1} = h\mathbf{T}h^{-1}$ are not involved. Apply Lemma 3.4 to $\tilde{H}$ and the subgroup $K = h\mathbf{T}h^{-1} \cap G$: this tells us that

$$\frac{|\tilde{H}^{(4)}|}{|\tilde{H}|} \geqslant \frac{|\tilde{H}^{(3)} \cap S|}{|\tilde{H}^{(2)} \cap S|}.$$

But since $L = \tilde{H}^{(2)}$ and $\mathbf{S}$ are not involved, we have $|\tilde{H}^{(2)} \cap S| \leqslant 2$, by the easy part of the Key Proposition 3.11, and therefore

$$\frac{|\tilde{H}^{(4)}|}{|\tilde{H}|} \geqslant \frac{1}{2}|\tilde{H}^{(3)} \cap S|.$$

However, L and $\mathbf{T}$ *are* involved. Now, for any $x = (h_1h_2)(h_3h_4) \in L \cap Tt$, we have

$$hxh^{-1} = (hh_1)(h_2h_3)(h_4h^{-1}) \in \tilde{H}^{(3)} \cap hTh^{-1} = \tilde{H}^{(3)} \cap S,$$

and hence

$$|\tilde{H}^{(3)} \cap S| \geqslant |\tilde{H}^{(2)} \cap T|.$$

By the second part of the Key Proposition 3.11 applied to $\tilde{H}$, we therefore obtain

$$\frac{|\tilde{H}^{(4)}|}{|\tilde{H}|} \geqslant 14^{-1}\tilde{\alpha}^{-4}|\tilde{H}|$$

with $\tilde{\alpha} = \operatorname{trp}(\tilde{H})$. Ruzsa's Lemma gives then

$$\alpha^6|H| \geqslant 14^{-1}\alpha^{-16}|H|^{4/3},$$

hence a rather stronger bound for α than before, namely

$$\alpha = \operatorname{trp}(H) \geqslant 14^{-1/22}|H|^{1/66} \geqslant 2^{-1/2}|H|^{1/66}. \tag{3.17}$$

To summarize, we have obtained three possible lower bounds of the right kind for α, namely (3.15), (3.16) and (3.17), one of which holds if $H^{(3)} \neq \mathrm{SL}_2(\mathbf{F}_p)$. All imply (3.14), and hence we are done.

3.4. **Diameter bound.** Corollary 1.3 is a well-known consequence of the growth theorem: by induction on $j \geqslant 1$, we see using Helfgott's Theorem that given a symmetric generating set $S \subset G = \mathrm{SL}_2(\mathbf{F}_p)$, either $\operatorname{diam} \mathcal{C}(G,S) \leqslant 3^j$, or

$$|H^{(3^j)}| \geqslant |H|^{(1+\delta)^j}$$

where $H = S \cup \{1\}$. For

$$j = \Big\lceil \frac{\log |G|}{\log(1+\delta)} \Big\rceil,$$

the second alternative is impossible, and hence

$$\operatorname{diam} \mathcal{C}(G,S) \leqslant 3^j \leqslant 3(\log |G|)^{1/\log(1+\delta)},$$

which gives the result since $1/\log(1+1/1344) \leqslant 1345$.

4. The Bourgain-Gamburd method

The method of Bourgain and Gamburd [1] leads, from Helfgott's growth theorem, to a proof that the Cayley graphs modulo primes of a Zariski-dense subgroup of $\mathrm{SL}_2(\mathbf{Z})$ form an expander family. Applying this method straightforwardly with explicit estimates (as done in [11, Ch. 4]), one obtains explicit expansion bounds (either for the spectral gap of the combinatorial Laplace operator, or for the discrete Cheeger constant). However, these constants are typically very small. We show in this section how a small modification of the original L^2-flattening lemma of Bourgain and Gamburd gives significant improvement (roughly speaking, exponentially better); such modifications are already described, e.g., by Varjú [20].

4.1. **The L^2-flattening inequality.** This section applies – in principle – to all finite groups, and the basic expansion criterion that we derive (Corollary 4.4, following essentially Bourgain and Gamburd) is also of independent interest.

In rough outline – and probabilistic language –, the idea is to show that if two independent $\mathrm{SL}_2(\mathbf{F}_p)$-valued symmetrically distributed random variables X_1 and X_2 are not too concentrated, but also not very uniformly distributed on $\mathrm{SL}_2(\mathbf{F}_p)$, then their product X_1X_2 will be significantly more uniformly distributed, *unless* there are obvious reasons why this should fail to hold. These exceptional possibilities can then be handled separately.

Applying this to some suitable step X_k of the random walk (where the initial condition is obtained by different means), this result leads to successive great improvements of the uniformity of the distribution for X_{2k}, X_{3k}, ..., X_{jk}, until the assumptions of the lemma fail. In that situation, the index $m = jk$ is of size about $\log |G|$, and $\mathbf{P}(X_{2m} = 1)$ gives a suitable upper-bound on the number of cycles to obtain expansion, by a variant of what might be called the Huxley-Sarnak-Xue method (see [9] and [17]), as we now recall.

For a finite group G, we denote my $d(G)$ the minimal dimension of a non-trivial irreducible unitary representation of G. Moreover, if X is a G-valued symmetrically-distributed random variable, we define the *return probability* $\mathrm{rp}(X)$ by

$$\mathrm{rp}(X) = \mathbf{P}(X_1X_2 = 1),$$

where (X_1, X_2) are independent random variables with the same distribution as X, or equivalently

$$\mathrm{rp}(X) = \sum_{g \in G} \mathbf{P}(X = g)^2.$$

If S is a symmetric generating subset of G and $\Gamma = \mathcal{C}(G,S)$ the associated Cayley graph, with random walk (starting at the origin) $(X_n)_{n\geqslant 0}$, we then have the upper-bounds

$$\varrho_\Gamma \leqslant \Bigl(\frac{|G|}{d(G)}\mathrm{rp}(X_m)\Bigr)^{1/(2m)} \tag{4.1}$$

for the spectral radius ϱ_Γ of the Markov operator of Γ, where $m \geqslant 1$ is arbitrary. (This is the "counting cycles" argument, with the input coming from the multiplicity of non-trivial eigenvalues.) Thus obtaining graphs where the right-hand side is bounded away from 1 is our goal.

We consider now two independent (not necessarily identically-distributed) G-valued random variables X_1, X_2 and let

$$\mathrm{rp}^+(X_1,X_2) = \max(\mathrm{rp}(X_1), \mathrm{rp}(X_2)).$$

We attempt to bound $\mathrm{rp}(X_1X_2)$ in terms of $\mathrm{rp}^+(X_1,X_2)$. To do this while still remaining at a level of great generality, the following definition will be useful:

Definition 4.1 (Flourishing)**.** For $\delta > 0$, a finite group G is δ-flourishing if any symmetric subset $H \subset G$, containing 1, which generates G and has tripling constant $\mathrm{trp}(H) < |H|^\delta$ satisfies $H^{(3)} = G$.

In particular, Theorem 1.2 states that all groups $\mathrm{SL}_2(\mathbf{F}_p)$, for p prime, are 1/1344-flourishing.

We will prove a general L^2-flattening theorem, which may be of general interest. In order to somehow streamline the proof, we do not explicitly describe here what "G large enough" means. However, all relevant steps where a condition on the size of G occurs are clearly marked, and in the second part of Section 4.3, we will look back to express these as explicit inequalities.

Theorem 4.2 (L^2-flattening conditions)**.** *Let G be a finite group which is δ-flourishing for some δ with $0 < \delta \leqslant 1$. Let X_1, X_2 be symmetric independent G-valued random variables.*

Let $0 < \gamma < 1$ be given, and assume that

$$\mathbf{P}(X_1 \in xH) \leqslant |G|^{-\gamma} \tag{4.2}$$

for all proper subgroups $H \subset G$ and all $x \in G$.

Then for any $\varepsilon > 0$, there exists $\delta_1 > 0$ and $c_3 > 0$, depending only on ε, δ and γ, such that

$$\mathrm{rp}(X_1X_2) \leqslant c_3 \max\Bigl\{\frac{1}{|G|^{1-\varepsilon}}, \frac{\mathrm{rp}^+(X_1,X_2)}{|G|^{\delta_1}}\Bigr\} \tag{4.3}$$

when $|G|$ is large enough in terms of $(\varepsilon,\delta,\gamma)$.

More precisely, one may take

$$\delta_1 = \frac{1}{2}\min\Bigl(\frac{\delta\gamma}{2c_2+1}, \frac{\varepsilon}{2c_2}\Bigr) \tag{4.4}$$

where $c_2 = 937$ is as in Theorem 2.1 and

$$c_3 \leqslant 2^{14}c_1^2 \leqslant 2^{3256}.$$

Proof. By definition, we have

$$\mathrm{rp}(X_1X_2) = \sum_{g\in G}\mathbf{P}(X_1X_2 = g)^2.$$

We now decompose the ranges of the distribution functions

$$\nu_i(x) = \mathbf{P}(X_i = x)$$

into dyadic intervals. Consider a parameter $I \geqslant 1$, to be chosen later, and decompose

$$[\min \mathbf{P}(X = x), \max \mathbf{P}(X = x)] \subset [0, 1] = \mathcal{I}_0 \cup \mathcal{I}_1 \cup \cdots \cup \mathcal{I}_I$$

where

$$\mathcal{I}_i = \begin{cases}]2^{-i-1}, 2^{-i}], & \text{for } 0 \leqslant i < I \\ [0, 2^{-I}] & \text{for } i = I. \end{cases}$$

This gives two partitions of G in subsets

$$A_{j,i} = \{x \in G \mid \nu_j(x) = \mathbf{P}(X_j = x) \in \mathcal{I}_i\},$$

for $j = 1$, 2. We note that

$$|A_{j,i}| \leqslant 2^{i+1} \tag{4.5}$$

for $j = 1$, 2 and $0 \leqslant i < I$.

Using this decomposition into the formula above, and the fact that

$$\mathbf{P}(X_1X_2 = g, X_1 \in A_{1,I} \text{ or } X_2 \in A_{2,I}) \leqslant \mathbf{P}(X_1 \in A_{1,I}) + \mathbf{P}(X_2 \in A_{2,I}) \leqslant \frac{|G|}{2^{I-1}},$$

we obtain

$$\begin{aligned} \mathrm{rp}(X_1X_2) &= \sum_{g\in G}\Big(\sum_{0\leqslant i,j\leqslant I} \mathbf{P}(X_1X_2 = g, X_1 \in A_{1,i},\ X_2 \in A_{2.j})\Big)^2 \\ &\leqslant 8|G|^3 2^{-2I} + 2\sum_{g\in G}\Big(\sum_{0\leqslant i,j<I} \mathbf{P}(X_1X_2 = g, X_1 \in A_{1,i},\ X_2 \in A_{2,j})\Big)^2 \\ &\leqslant 2^{3-2I}|G|^3 + 2I^2 \sum_{0\leqslant i,j<I}\sum_{g\in G} \mathbf{P}(X_1X_2 = g, X_1 \in A_{1,i},\ X_2 \in A_{2,j})^2 \end{aligned}$$

by the Cauchy-Schwarz inequality. Furthermore, the inner sum (say, $B(A_{1,i}, A_{2,j})$) in the second term is given by

$$\begin{aligned} B(A_{1,i}, A_{2,j}) &= \sum_{g\in G} \mathbf{P}(X_1X_2 = g,\ X_1 \in A_{1,i},\ X_2 \in A_{2,j})^2 \\ &= \sum_{g\in G}\Big(\sum_{\substack{(x,y)\in A_{1,i}\times A_{2,j}\\ xy=g}} \mathbf{P}(X_1 = x)\,\mathbf{P}(X_2 = y)\Big)^2 \\ &= \sum_{\substack{x_1,x_2\in A_{1,i},y_1,y_2\in A_{2,j}\\ x_1y_1=x_2y_2}} \nu_1(x_1)\nu_1(x_2)\nu_2(y_1)\nu_2(y_2) \\ &\leqslant 2^{-2i-2j}|\{(x_1, x_2, y_1, y_2) \in A_{1,i}^2 \times A_{2,j}^2 \mid x_1y_1 = x_2y_2\}| \\ &= 2^{-2i-2j}E(A_{1,i}, E_{2,j}) \end{aligned}$$

where $E(A, B)$ denotes the multiplicative energy.

Thus we have proved that

$$\mathrm{rp}(X_1X_2) \leqslant 2^{3-2I}|G|^3 + 2I^2 \sum_{0\leqslant i,j<I} 2^{-2(i+j)}E(A_{1,i}, A_{2,j}). \tag{4.6}$$

We now want to get upper-bounds in terms of the return probability $\mathrm{rp}^+(X_1, X_2)$. This is done in different ways, depending on the size of the subsets $A_{1,i}$, $A_{2,j}$. We recall first the "trivial" bounds

$$E(A, B) \leqslant \min(|A|^2|B|, |A||B|^2). \tag{4.7}$$

We claim that for all i and j, we have

$$2^{-2(i+j)}E(A_{1,i}, A_{2,j}) \leqslant 2^4\mathrm{rp}^+(X_1, X_2)e(A_{1,i}, A_{2,j}), \tag{4.8}$$

and that, for all $\alpha \geqslant 1$, we have

$$2^{-2(i+j)}E(A_{1,i}, A_{2,j}) \leqslant \alpha^{-1}\mathrm{rp}(X) \tag{4.9}$$

unless

$$\frac{|A_{1,i}|}{2^i} \geqslant \frac{1}{2\sqrt{\alpha}}, \qquad \frac{|A_{2,j}|}{2^j} \geqslant \frac{1}{2\sqrt{\alpha}}. \tag{4.10}$$

To see (4.8), we remark that

$$\begin{aligned}\mathrm{rp}^+(X_1, X_2) \geqslant \frac{1}{2}(\mathrm{rp}(X_1) + \mathrm{rp}(X_2)) &= \frac{1}{2}\sum_{g\in G}(\mathbf{P}(X_1 = g)^2 + \mathbf{P}(X_2 = g)^2)\\ &\geqslant \frac{1}{2}\Bigl(\frac{|A_{1,i}|}{2^{2+2i}} + \frac{|A_{2,j}|}{2^{2+2j}}\Bigr) \geqslant \frac{1}{4}\frac{(|A_{1,i}||A_{2,j}|)^{1/2}}{2^{i+j}}.\end{aligned}$$

for any choice of i and j. Hence we get

$$\begin{aligned}2^{-2(i+j)}E(A_{1,i}, A_{2,j}) &= 2^{-2(i+j)}e(A_{1,i}, A_{2,j})(|A_{1,i}||A_{2,j}|)^{3/2}\\ &\leqslant 4\mathrm{rp}^+(X_1, X_2)e(A_{1,i}, A_{2,j})\frac{|A_{1,i}||A_{2,j}|}{2^{i+j}}\\ &\leqslant 16\mathrm{rp}^+(X_1, X_2)e(A_{1,i}, A_{2,j})\end{aligned}$$

by (4.5).

As for (4.9), if we assume that $2^{-2(i+j)}E(A_{1,i}, A_{2,j}) > \alpha^{-1}\mathrm{rp}^+(X_1, X_2)$, then we write simply

$$2^{-2(i+j)}|A_{1,i}|^2|A_{2,j}| \geqslant 2^{-2(i+j)}E(A_{1,i}, A_{2,j}) \geqslant \alpha^{-1}\frac{|A_{2,j}|}{2^{2+2j}},$$

using (4.7), and get the first inequality of (4.10), the second being obtained symmetrically.

With these results, we now fix some parameter $\alpha \geqslant 1$, and let

$$P_\alpha = \{(i, j) \mid 0 \leqslant i, j < I, \quad |A_{1,i}| \geqslant 2^{i-1}\alpha^{-1} \text{ and } |A_{2,j}| \geqslant 2^{j-1}\alpha^{-1}\}.$$

For $(i, j) \notin P_\alpha$, we have

$$2^{-2(i+j)}E(A_{1,i}, A_{2,j}) \leqslant \alpha^{-2}\mathrm{rp}^+(X_1, X_2)$$

by (4.9) and (4.10), and thus from (4.6), we have shown that

$$\mathrm{rp}(X_1X_2) \leqslant 2^{3-2I}|G|^3 + 2\alpha^{-2}\mathrm{rp}^+(X_1, X_2)I^4 + 32\mathrm{rp}^+(X_1, X_2)I^2 \sum_{(i,j)\in P_\alpha} e(A_{1,i}, A_{2,j})$$

(estimating the size of the complement of P_α by I^2). We select

$$I = \Bigl\lceil \frac{2\log 2|G|}{\log 2}\Bigr\rceil \leqslant 3\log(3|G|),$$

and hence obtain

$$\mathrm{rp}(X_1X_2) \leqslant \frac{1}{|G|} + 2^8\mathrm{rp}^+(X_1, X_2)(\log 3|G|)^2\Bigl\{\frac{(\log 3|G|)^2}{\alpha^2} + 2\sum_{(i,j)\in P_\alpha} e(A_{1,i}, A_{2,j})\Bigr\}.$$

We apply this bound with $\alpha = |G|^{\delta_0}$, where $\delta_0 > 0$ will be chosen later. Thus

$$\mathrm{rp}(X_1X_2) \leqslant \frac{1}{|G|} + 2^8 \mathrm{rp}^+(X_1, X_2)(\log 3|G|)^4 |G|^{-2\delta_0} + \\ 2^9 (\log 3|G|)^2 \mathrm{rp}^+(X_1, X_2) \sum_{(i,j)\in P_\alpha} e(A_{1,i}, A_{2,j}).$$

Let then

$$R_\alpha = \{(i, j) \in P_\alpha \mid e(A_{1,i}, A_{2,j}) \geqslant \alpha^{-1}\} \subset P_\alpha,$$

so that the contribution of those $(i, j) \in P_\alpha$ which are not in R_α, together with the middle term, can be bounded by

$$\frac{2^{13}(\log 3|G|)^4}{|G|^{\delta_0}} \mathrm{rp}^+(X_1, X_2).$$

We can now analyze the set R_α; it turns out to be very restricted when δ_0 is chosen small enough. By Theorem 2.1, for each $(i, j) \in R_\alpha$, there exists a β_1-approximate subgroup $\mathsf{H}_{i,j}$ and elements $(x_i, y_j) \in A_{1,i} \times A_{2,j}$ such that

$$|\mathsf{H}_{i,j}| \leqslant \beta_2 |A_{1,i}|, \qquad |A_{1,i} \cap x_i \mathsf{H}_{i,j}| \geqslant \beta_3^{-1} |A_{1,i}|, \qquad |A_{2,j} \cap \mathsf{H}_{i,j} y_j| \geqslant \beta_3^{-1} |A_{2,j}|,$$

and with tripling constant bounded by β_4, where the β_i are bounded qualitatively by

$$\beta_i \leqslant c_1 |G|^{c_2\delta_0}$$

for some absolute constants, which we take to be $c_1 = 2^{2112}$, $c_2 = 937$ using (2.2). We then note first that if $H_{i,j}$ denotes the "ordinary" subgroup generated by $\mathsf{H}_{i,j}$, we have

$$\mathbf{P}(X_1 \in x_i H_{i,j}) \geqslant \mathbf{P}(X_1 \in x_i \mathsf{H}_{i,j}) \geqslant \\ \mathbf{P}(X_1 \in A_{1,i} \cap x_i \mathsf{H}_{i,j}) \geqslant \frac{1}{\beta_3} \frac{|A_{1,i}|}{2^{i+1}} \geqslant \frac{1}{4\beta_3\alpha} \geqslant \frac{1}{4c_1|G|^{(1+c_2)\delta_0}}, \tag{4.11}$$

where we used the definition of P_α. If δ_0 is small enough that

$$(1 + c_2)\delta_0 < \gamma, \tag{4.12}$$

and if $|G|$ is large enough, this is not compatible with (4.2), and we can therefore assume that each $\mathsf{H}_{i,j}$ (if any!) generates the group G.

We next observe that $\mathsf{H}_{i,j}$ can not be extremely small. Indeed, we have

$$|\mathsf{H}_{i,j}| \geqslant |x_i \mathsf{H}_{i,j} \cap A_{1,i}| \geqslant \beta_3^{-1} |A_{1,i}|,$$

on the one hand, and by applying (4.2) with $H = 1$, we can see that $A_{1,i}$ is not too small, namely

$$|A_{1,i}| \geqslant \frac{\mathbf{P}(X \in A_{1,i})}{\max_{g\in G} \mathbf{P}(X_1 = g)} \geqslant |G|^\gamma \, \mathbf{P}(X_1 \in A_{1,i}) \geqslant \frac{|G|^\gamma |A_{1,i}|}{2^{i+1}} \geqslant \frac{|G|^\gamma}{4\alpha}$$

using again the definition of P_α.

This gives the lower bound

$$|\mathsf{H}_{i,j}| \geqslant \frac{|G|^\gamma}{4\alpha\beta_3} \geqslant \frac{1}{4c_1} |G|^{\gamma_1} \tag{4.13}$$

with $\gamma_1 = \gamma - \delta_0(1 + c_2)$ (which is > 0 by (4.12)), and then leads to control of the tripling constant, namely

$$\mathrm{trp}(\mathsf{H}_{i,j}) \leqslant \beta_4 \leqslant c_1 |G|^{c_2\delta_0} \leqslant c_1 (4c_1)^{\gamma_1^{-1}} |\mathsf{H}_{i,j}|^{c_2\delta_0\gamma_1^{-1}}. \tag{4.14}$$

Since we assumed that G is δ-flourishing, we see from Definition 4.1 that if δ_0 is such that

$$\frac{c_2\delta_0}{\gamma_1} = \frac{c_2\delta_0}{\gamma - (1+c_2)\delta_0} < \delta, \tag{4.15}$$

and again if $|G|$ is large enough, the approximate subgroup $\mathsf{H}_{i,j}$ must in fact be very large, specifically it must satisfy

$$\mathsf{H}_{i,j} \cdot \mathsf{H}_{i,j} \cdot \mathsf{H}_{i,j} = G,$$

and in particular

$$|\mathsf{H}_{i,j}| \geqslant \frac{|G|}{\beta_4} \geqslant \frac{1}{c_1}|G|^{1-c_2\delta_0}.$$

Intuitively, this implies that X_1 and X_2 are already rather uniformly distributed over G, and hence that $\mathrm{rp}^+(X_1, X_2)$ is already too small to be significantly improved at the level of X_1X_2. To express this idea concretely, we go back to the first stage of the argument, namely (4.6): the contribution to $\mathrm{rp}(X_1X_2)$ coming from (i,j) was bounded by

$$2^{-2(i+j)}E(A_{1,i}, A_{2,j}) \leqslant \frac{|A_{1,i}||A_{2,j}|^2}{2^{2(i+j)}} \leqslant \frac{1}{2^{i-3}}$$

by (4.5). But then we also have

$$2^{i+1} \geqslant |A_{1,i}| \geqslant |A_{1,i} \cap x_i\mathsf{H}_{i,j}| \geqslant \frac{|A_{1,i}|}{\beta_3} \geqslant \frac{|\mathsf{H}_{i,j}|}{\beta_2\beta_3} \geqslant \frac{|G|}{\beta_2\beta_3\beta_4} \geqslant c_1^{-2}|G|^{1-2c_2\delta_0},$$

(as one sees from (2.2) that $\beta_2\beta_3\beta_4 \leqslant c_1^2\alpha^{2c_2} \leqslant c_1^2|G|^{2c_2\delta_0}$) and therefore

$$2^{-2(i+j)}E(A_{1,i}, A_{2,j}) \leqslant 16c_1^2|G|^{-1+2c_2\delta_0}.$$

Using again the trivial bound $I^2 \leqslant 9(\log 3|G|)^2$ for the number of possible pairs (i,j) to which this applies, the conclusion is an inequality

$$\mathrm{rp}(X_1X_2) \leqslant 2^{11}c_1^2\frac{(\log 3|G|)^4}{|G|^{1-2c_2\delta_0}} + 2^{13}\frac{(\log 3|G|)^4}{|G|^{\delta_0}}\mathrm{rp}^+(X_1, X_2), \tag{4.16}$$

which holds (under the assumptions that $|G|$ is sufficiently large) for all δ_0 small enough so that (4.12) and (4.15) are satisfied. It is elementary that (4.15) is stronger than (4.12) and is equivalent with

$$\delta_0 < \frac{\delta\gamma}{(1+\delta)c_2 + \delta},$$

which holds when $\delta_0 < \delta\gamma/(2c_2+1)$ (since we assume $\delta \leqslant 1$).

Thus we can apply this for

$$\delta_0 = \min\Big(\frac{\delta\gamma}{2c_2+1}, \frac{\varepsilon}{2c_2}\Big) = 2\delta_1,$$

where δ_1 is given by (4.4). Then for $|G|$ large enough, (4.16) implies (4.3), and hence we have finished the proof of Theorem 4.2. □

In order to apply this theorem iteratively, we need also the following simple observation of "increase of uniformity".

Lemma 4.3 (Uniformity can only increase)**.** *Let G be a finite group, S a symmetric generating subset, and let (X_n) be the corresponding random walk on $\mathcal{C}(G,S)$. For any $n \geqslant 1$ and $m \geqslant n$, we have* $\mathrm{rp}(X_m) \leqslant \mathrm{rp}(X_n)$.

Proof. By the spectral interpretation of the return probability, we have

$$\mathrm{rp}(X_m) = \mathbf{P}(X_{2m} = 1) = \frac{1}{|G|}\,\mathrm{Tr}(M^{2m}), \qquad \mathrm{rp}(X_{2n}) = \mathbf{P}(X_{2n} = 1) = \frac{1}{|G|}\,\mathrm{Tr}(M^{2n})$$

were M is the Markov operator. Since all eigenvalues of M^2 are non-negative and $\leqslant 1$, it follows that

$$\mathrm{Tr}(M^{2m}) \leqslant \mathrm{Tr}(M^{2n})$$

for $m \geqslant n$, as desired. ☐

We can summarize all this as follows (with the same remark as before concerning our handling of the conditions on the size of G):

Corollary 4.4 (The Bourgain-Gamburd expansion criterion)**.** *Let $\boldsymbol{c} = (c, d, \delta, \gamma)$ be a tuple of positive real numbers, and let $\mathcal{G}(\boldsymbol{c})$ be the family of all finite connected Cayley graphs $\mathcal{C}(G, S)$ for which the following conditions hold:*

(1) *We have $d(G) \geqslant |G|^d$;*
(2) *The group G is δ-flourishing;*
(3) *For the random walk (X_n) on G with $X_0 = 1$, we have that*

$$\mathbf{P}(X_{2k} \in xH) \leqslant |G|^{-\gamma}$$

for some $k \leqslant c \log |G|$ and all $x \in G$ and proper subgroups $H \subset G$.

Then, for any $\Gamma \in \mathcal{G}(\boldsymbol{c})$ with $|\Gamma|$ large enough, the spectral gap of the normalized Laplace operator of Γ satisfies

$$\lambda_1(\Gamma) \geqslant 1 - \exp\Bigl(-\frac{d}{4cj}\Bigr),$$

where

$$j \leqslant 8 \max\Bigl(\frac{2c_2 + 1}{\delta\gamma}, \frac{16c_2}{7d}\Bigr).$$

Note that it is not clear at this point that this corollary is not an empty statement (or one that applies at most to finitely many graphs with a bounded valency). But in the next section we will check that it applies to the situation of Theorem 1.1 to prove that certains families of Cayley graphs are expanders.

Proof. Let $\Gamma = \mathcal{C}(G, S)$ be a graph in $\mathcal{G}(\boldsymbol{c})$. We will apply Theorem 4.2 with $\varepsilon = 7d/8$ so that

$$\delta_1 = \frac{1}{2} \min\Bigl(\frac{\delta\gamma}{2c_2 + 1}, \frac{7d}{16c_2}\Bigr)$$

When $|G|$ is large enough, we can rephrase the conclusion using the simpler inequality

$$\mathrm{rp}(Y_1Y_2) \leqslant c_3 \max\Bigl(\frac{1}{|G|^{1-7d/8}}, \frac{\mathrm{rp}^+(Y_1, Y_2)}{|G|^{\delta_1}}\Bigr) \leqslant \max\Bigl(\frac{1}{|G|^{1-3d/4}}, \frac{\mathrm{rp}^+(Y_1, Y_2)}{|G|^{\delta_1/2}}\Bigr), \tag{4.17}$$

for random variables Y_1, Y_2 which satisfy the assumptions of this theorem.

We take $Y_1 = X_k$ where $k = \lfloor c \log |G| \rfloor$ is given by (3), and $Y_2 = X_k^{-1} X_{jk}$ for $j \geqslant 2$. These are indeed independent and symmetric random variables. Conditions (2) and (3) imply that we can apply Theorem 4.2 to these random variables for any $j \geqslant 2$. Since Y_2 is distributed like $X_{(j-1)k}$, we have

$$\mathrm{rp}^+(Y_1, Y_2) \leqslant \mathrm{rp}(Y_1) = \mathrm{rp}(X_k)$$

by the last lemma. Thus, applying the theorem, we obtain by induction

$$\mathrm{rp}(X_{jk}) \leqslant \mathrm{rp}(X_k)|G|^{-j\delta_1/2} \leqslant |G|^{-j\delta_1/2}$$

when j is such that

$$|G|^{1-3d/4} > |G|^{j\delta_1/2},$$

and for larger j, we get

$$\mathrm{rp}(X_{jk}) \leqslant |G|^{-1+3d/4}.$$

In particular, we obtain this last inequality for

$$j = \Big\lceil \frac{2(1-3d/4)}{\delta_1} \Big\rceil \leqslant \frac{4}{\delta_1} \leqslant 8 \max\Big(\frac{2c_2+1}{\delta\gamma}, \frac{16c_2}{7d}\Big),$$

which, by the "cycle-counting" inequality (4.1), gives

$$\varrho_\Gamma \leqslant (|G|^{1-d}\mathrm{rp}(X_{jk}))^{1/(2jk)} \leqslant \exp\Big(-\frac{d}{4jc}\Big),$$

which thus proves the theorem. □

4.2. **Expansion bounds for** SL_2**.** Theorem 1.1 will now be proved by applying the criterion of Corollary 4.4. Thus we will consider the groups $G_p = \mathrm{SL}_2(\mathbf{F}_p)$ for p prime, for which Condition (1) of the Bourgain-Gamburd criterion (which is purely a group-theoretic property) is given by

$$d(\mathrm{SL}_2(\mathbf{F}_p)) = \frac{p-1}{2}$$

for $p \geqslant 3$ (a result of Frobenius), which gives a value of d arbitrarily close to $1/3$, for p large enough. Condition (2) is given by Helfgott's Theorem, with $\delta = 1/1344$. Note that it is purely a property of the groups $\mathrm{SL}_2(\mathbf{F}_p)$.

Condition (3), on the other hand, depends on the choice of generating sets. The symmetric generating sets S_p in Theorem 1.1 are assumed to be obtained by reduction modulo p of a fixed symmetric subset $S \subset \mathrm{SL}_2(\mathbf{Z})$. We will argue first under the additional assumption that $S \subset \mathrm{SL}_2(\mathbf{Z})$ generates a free group.

We begin with a classical proposition, whose idea goes back to Margulis. For the statement, recall that the norm of a matrix $g \in \mathrm{GL}_n(\mathbf{C})$ is defined by

$$\|g\| = \max_{v\neq 0} \frac{|\langle gv, w\rangle|}{\|v\|\|w\|}$$

where $\langle \cdot, \cdot \rangle$ is the standard inner product on $\mathbf{C}^n$. This satisfies

$$\|g_1g_2\| \leqslant \|g_1\|\|g_2\|, \qquad \max_{i,j} |g_{i,j}| \leqslant \|g\| \text{ for } g = (g_{i,j}), \tag{4.18}$$

the latter because $g_{i,j} = \langle ge_i, e_j \rangle$ in terms of the canonical basis.

Proposition 4.5 (Large girth for finite Cayley graphs)**.** *Let $S \subset \mathrm{SL}_2(\mathbf{Z})$ be a symmetric set, and let $\Gamma = \mathcal{C}(G,S)$ be the corresponding Cayley graphs. Let $\tau > 0$ be defined by*

$$\tau^{-1} = \log \max_{s\in S} \|s\| > 0, \tag{4.19}$$

which depends only on S.

(1) *For all primes p and all $r < \tau \log(p/2)$, where $G_p = \mathrm{SL}_2(\mathbf{F}_p)$, the subgraph Γ_r induced by the ball of radius r in Γ maps injectively to $\mathcal{C}(G_p, S)$.*

(2) *If G is freely generated by S, in particular $1 \notin S$, the Cayley graph $\mathcal{C}(G_p, S)$ contains no cycle of length $< 2\tau\log(p/2)$, i.e., its girth* $\mathrm{girth}(\mathcal{C}(\mathrm{G_p},\mathrm{S}))$ *is at least equal to $2\tau\log(p/2)$.*

Proof. The main point is that if all coordinates of two matrices g_1, $g_2 \in \mathrm{SL}_2(\mathbf{Z})$ are less than $p/2$ in absolute value, a congruence $g_1 \equiv g_2 \pmod p$ is equivalent to the equality $g_1 = g_2$. And because G is freely generated by S, knowing a matrix in G is equivalent to knowing its expression as a word in the generators in S.

Thus, let x be an element in the ball of radius r centered at the origin. By definition, x can be expressed as

$$x = s_1 \cdots s_m$$

with $m \leqslant r$ and $s_i \in S$. Using (4.18), we get

$$\max_{i,j} |x_{i,j}| \leqslant \|x\| \leqslant \|s_1\| \cdots \|s_m\| \leqslant e^{m/\tau} \leqslant e^{r/\tau}.$$

Applying the beginning remark and this fact to two elements x and y in the ball $\mathcal{B}_1(r)$ of radius r centered at 1, for r such that $e^{r/\tau} < \frac{p}{2}$, it follows that $x \equiv y \pmod p$ implies $x = y$, which is (1).

Then (2) follows because any embedding of a cycle $\gamma : C_m \to \mathcal{C}(G_p, S)$ such that $\gamma(0) = 1$ and such that

$$d(1, \gamma(i)) \leqslant m/2 < \tau \log(p/2)$$

for all i can be lifted to the cycle (of the same length) with image in the Cayley graph of G with respect to S, and if S generates freely G, the latter graph is a tree. Thus a cycle of length $m = \mathrm{girth}(\mathcal{C}(\mathrm{G_p}, \mathrm{S}))$ must satisfy $m/2 \geqslant \tau \log(p/2)$. □

We can now check Condition (3) in the Bourgain-Gamburd criterion, first for cosets of the trivial subgroup, i.e., for the probability that X_n be a fixed element when n is of size $c \log p$ for some fixed (but small) $c > 0$. As we did earlier, we clearly mark where we impose conditions on the size of p, and these will be made explicit in Section 4.3.

Corollary 4.6 (Decay of probabilities)**.** *Let $S \subset \mathrm{SL}_2(\mathbf{Z})$ be a symmetric set, G the subgroup generated by S. Assume that S freely generates G. Let p be a prime such that the reduction S_p of S modulo p generates $G_p = \mathrm{SL}_2(\mathbf{F}_p)$, and let (X_n) be the random walk on $\mathcal{C}(G_p, S_p)$ with $X_0 = 1$. Let*

$$\tau^{-1} = \log \max_{s \in S} \|s\| > 0,$$

as in Proposition 4.5.

Fix a constant c with $0 < c \leqslant 1$. If p is large enough, depending on c and S, then for

$$n = c\lfloor \tau \log(p/2) \rfloor$$

and any $x \in \mathrm{SL}_2(\mathbf{F}_p)$, we have

$$\mathbf{P}(X_n = x) \leqslant |G_p|^{-c\gamma_1} \tag{4.20}$$

where

$$\gamma_1 = \frac{\tau(\log |S|)}{16}. \tag{4.21}$$

More precisely, this holds for all

$$p \geqslant \max\Bigl(17, 2\exp\Bigl(\frac{c\tau(\log |S|)^2}{8}\Bigr)\Bigr). \tag{4.22}$$

The "extra" parameter c will be useful in the argument involving all proper subgroups H below.

Proof. There exists $\tilde{x} \in G$ such that $\tilde{x}$ reduces to x modulo p and $\tilde{x}$ is at the same distance to 1 as x, and by Proposition 4.5, (2), we have

$$\mathbf{P}(X_n = x) = \mathbf{P}(\tilde{X}_n = \tilde{x}),$$

for $n \leqslant \tau \log(p/2)$, where $(\tilde{X}_n)$ is the random walk starting at 1 on the $|S|$-regular tree $\mathcal{C}(G, S)$. By a well-known result of Kesten, we have

$$\mathbf{P}(\tilde{X}_n = \tilde{x}) \leqslant r^{-n} \quad \text{with} \quad r = \frac{|S|}{2\sqrt{|S|-1}},$$

for all $n \geqslant 1$ and all $\tilde{x} \in G$. Since $c \leqslant 1$ we have

$$n = c\lfloor c\tau \log(p/2)\rfloor \geqslant c\tau \log(p/2) - 1,$$

and we obtain

$$\mathbf{P}(X_n = x) \leqslant r\Big(\frac{p}{2}\Big)^{-c\tau \log r} \leqslant \Big(\frac{p}{2}\Big)^{-\frac{1}{2}c\tau \log r},$$

for $p \geqslant 2r^{2/(c\tau \log r)}$. Using the inequality

$$\frac{p}{2} \geqslant |G_p|^{1/4}$$

for $p \geqslant 17$, this becomes

$$\mathbf{P}(X_n = x) \leqslant |G_p|^{-c\tau(\log r)/8}$$

for all $p \geqslant \max(17, 2r^{1/(c\tau \log r)})$. Since $r \leqslant \sqrt{|S|}$, we have

$$\frac{c\tau(\log r)}{8} \geqslant \frac{c\tau(\log |S|)}{16} = c\gamma_1,$$

which gives the desired result. □

In order to deal with cosets of other proper subgroups of $\mathrm{SL}_2(\mathbf{F}_p)$, we will exploit the fact that those subgroups are very well understood, and in particular, there is no proper subgroup that is "both big and complicated". Precisely, by results going back to Dickson, one knows that for $p \geqslant 5$, if $H \subset \mathrm{SL}_2(\mathbf{F}_p)$ is a proper subgroup, one of the following two properties holds:

(1) The order of H is at most 120;

(2) For all $(x_1, x_2, x_3, x_4) \in H$, we have

$$[[x_1, x_2], [x_3, x_4]] = 1. \tag{4.23}$$

The first ones are "small", and will be easy to handle using (4.20). The second are, from the group-theoretic point of view, not very complicated (their commutator subgroups are abelian). The following *ad-hoc* lemma[8] takes care of them:

Proposition 4.7. *Let $k \geqslant 2$ be an integer and let $W \subset F_k$ be a subset of the free group on k generators $(a_1, \ldots, a_k)$ such that*

$$[[x_1, x_2], [x_3, x_4]] = 1 \tag{4.24}$$

for all $(x_1, x_2, x_3, x_4) \in W$. Then for any $m \geqslant 1$, we have

$$|\{x \in W \mid d_T(1, x) \leqslant m\}| \leqslant (4m+1)(8m+1) \leqslant 45m^2,$$

where T is the $|S|$-regular tree $\mathcal{C}(F_k, S)$, $S = \{a_i^{\pm 1}\}$.

[8] Note that this is the only place where using prime fields $\mathbf{F}_p$ instead of arbitrary finite fields really simplifies the argument, since (4.23) does not hold for proper subgroups of, say, $\mathrm{SL}_2(\mathbf{F}_{p^2})$.

Proof. The basic fact we need is that the condition $[x, y] = 1$ is very restrictive in F_k: precisely, for a fixed $x \neq 1$, we have $[x, y] = 1$ if and only if $y \in C_{F_k}(x)$, which is an infinite cyclic group. Denoting a generator by z, we find

$$(4.25)\qquad |\{y \in \mathcal{B}_1(m) \mid [x, y] = 1\}| = |\{h \in \mathbf{Z} \mid d_{T_k}(1, z^h) \leqslant m\}| \leqslant 2m + 1$$

since (a standard fact in free groups) we have $d_T(1, z^h) \geqslant |h|$.

Let W be a set satisfying the assumption (4.24), which we assume to be not reduced to $\{1\}$. We denote $W_m = W \cap \mathcal{B}_1(m)$. First, if $[x, y] = 1$ for all x, $y \in W_m$, then by taking a fixed $x \neq 1$ in W_m, we get $W_m \subset C_{F_k}(x) \cap \mathcal{B}_1(m)$, and (4.25) gives the result.

Otherwise, fix x_0 and y_0 in W_m such that $a = [x_0, y_0] \neq 1$. Then, for all y in W_m we have $[a, [x_0, y]] = 1$. Noting that $d_T(1, [x_0, y]) \leqslant 4m$, it follows again from the above that the number of possible values of $[x_0, y]$ is at most $8m + 1$ for $y \in W_m$.

Now for one such value $b = [x_0, y]$, we consider how many $y_1 \in W_m$ may satisfy $[x_0, y_1] = b$. We have $[x_0, y] = [x_0, y_1]$ if and only if $\varphi(yy_1^{-1}) = yy_1^{-1}$, where $\varphi(y) = x_0 y x_0^{-1}$ denotes the inner automorphism of conjugation by x_0. Hence y_1 satisfies $[x_0, y_1] = b$ if and only if $\varphi(yy_1^{-1}) = yy_1^{-1}$, which is equivalent to $yy_1^{-1} \in C_{F_k}(x_0)$. Taking a generator z of this centralizer again (note $x_0 \neq 1$), we get

$$\begin{aligned} |\{y_1 \in \mathcal{B}_1(m) \mid [x_0, y_1] = [x_0, y]\}| &= |\{h \in \mathbf{Z} \mid yz^h \in \mathcal{B}_1(m)\}| \\ &\leqslant |\{h \in \mathbf{Z} \mid z^h \in \mathcal{B}_1(2m)\}| \leqslant 4m + 1, \end{aligned}$$

since

$$d_T(1, z^h) = d_T(y, yz^h) \leqslant d_T(1, y) + d_T(1, yz^h) \leqslant 2m$$

for $h \in \mathbf{Z}$ such that yz^h is in $\mathcal{B}_1(m)$.

Hence we have $|W_m| \leqslant (4m + 1)(8m + 1)$ in that case, which proves the result. ∎

Using Corollary 4.6, we finally verify fully Condition (3) in Corollary 4.4:

Corollary 4.8 (Decay of probabilities, II)**.** *Let $S \subset \mathrm{SL}_2(\mathbf{Z})$ be a symmetric set, G the subgroup generated by S. Assume that S freely generates G. Let p be a prime such that the reduction S_p of S modulo p generates $G_p = \mathrm{SL}_2(\mathbf{F}_p)$, and let (X_n) be the random walk on $\mathcal{C}(G_p, S_p)$ with $X_0 = 1$. Let*

$$\tau^{-1} = \log \max_{s \in S} \|s\| > 0,$$

as in Proposition 4.5.

If p is large enough, then for

$$n = \Big\lfloor \frac{\tau}{32} \log(p/2) \Big\rfloor,$$

any $x \in \mathrm{SL}_2(\mathbf{F}_p)$ and any proper subgroup $H \subset \mathrm{SL}_2(\mathbf{F}_p)$, we have

$$(4.26)\qquad \mathbf{P}(X_n \in xH) \leqslant |G_p|^{-\gamma}$$

where

$$(4.27)\qquad \gamma = \frac{\tau(\log |S|)}{2^{10}}.$$

Proof. We start by noting that

$$\mathbf{P}(X_n \in xH)^2 \leqslant \mathbf{P}(X_{2n} \in H)$$

for all $x \in G_p$ and all subgroups $H \subset G_p$.

Consider first the case where (4.23) holds for H. Let $\tilde{H} \subset G$ be the pre-image of H under reduction modulo p. If $2n \leqslant \tau \log(p/2)$, then as in the proof of Corollary 4.6, we get

$$\mathbf{P}(X_{2n} \in H) = \mathbf{P}(\tilde{X}_{2n} \in \tilde{H}).$$

Provided n also satisfies the stronger condition $n \leqslant m = \frac{1}{16}\tau \log(p/2)$, any commutator

$$[[x_1, x_2], [x_3, x_4]]$$

with $x_i \in \tilde{H} \cap \mathcal{B}_1(n)$ is an element at distance at most $\tau \log(p/2)$ from 1 in the tree $\mathcal{C}(G, S)$, which reduces to the identity modulo p by (4.23), and therefore must be itself equal to 1. In other words, we can apply Proposition 4.7 to $W = \tilde{H} \cap \mathcal{B}_1(m)$ to deduce the upper bound

$$|\tilde{H} \cap \mathcal{B}_1(m)| \leqslant 45m^2.$$

We now take

$$n = \frac{1}{32}\lfloor \tau \log(p/2) \rfloor,$$

and we derive

$$\mathbf{P}(X_{2n} \in H) \leqslant |\tilde{H} \cap \mathcal{B}_1(m)| r^{-2n} \leqslant 45m^2 |G_p|^{-\gamma_1/16}$$

(where γ_1 is given by (4.21), as in Corollary 4.6), and hence

$$\mathbf{P}(X_n \in xH) \leqslant \frac{\sqrt{45}}{16}\tau(\log p/2)|G_p|^{-\gamma_1/32} \leqslant |G_p|^{-\gamma_1/64} \tag{4.28}$$

provided p is large enough, which is the conclusion in that case.

On the other hand, if (4.23) does not hold, we have $|H| \leqslant 120$, and for the same value of n we get

$$\mathbf{P}(X_n \in xH) \leqslant 120|G_p|^{-\gamma_1/32} \leqslant |G_p|^{-\gamma_1/64} \tag{4.29}$$

for p large enough, by Corollary 4.6 with $c = 1/32$. This gives again the desired result. □

4.3. **Summary.** We can now summarize how to obtain an explicit spectral gap, for large enough p, in the situation of Theorem 1.1, finishing the proof. We then explain how to quantify the condition on p.

We first consider the case where $S \subset \mathrm{SL}_2(\mathbf{Z})$ freely generates a free group of rank $\geqslant 2$ (in which case it is automatically Zariski-dense in SL_2).

<u>Step 1 (when p is large enough).</u> We have

$$d(G_p) = \frac{p-1}{2}$$

for $p \geqslant 3$. In particular, $d(G_p) \geqslant |G_p|^d$ for any $d < 1/3$ provided p is large enough in terms of d. Moreover, by Theorem 1.2, those groups are δ-flourishing with $\delta = 1/1344$.

For the random walk (X_n) on G_p associated to the generating set S_p, with $X_0 = 1$, we have

$$\mathbf{P}(X_{2k} \in xH) \leqslant |G|^{-\gamma}$$

when

$$k = \left\lfloor \frac{\tau}{32} \log(p/2) \right\rfloor \leqslant \frac{\tau}{96} \log(|G_p|)$$

with

$$\tau^{-1} = \log \max_{s \in S} \|s\|, \qquad \gamma = \frac{\tau(\log |S|)}{2^{10}}.$$

by (4.19) and (4.27). Thus in Corollary 4.4, we can take $c = 1/96$. The number of times we apply the basic L^2-flattening inequality is bounded by

$$j \leqslant 8 \max\Big(\frac{2c_2+1}{\delta\gamma}, \frac{8c_2}{d}\Big) \leqslant 8 \max\Big(\frac{1875 \cdot 1344}{\gamma}, 2500\Big),$$

and the spectral gap satisfies

$$\lambda_1(\Gamma) \geqslant 1 - \exp\Big(-\frac{d}{4cj}\Big),$$

for all p large enough, which is about

$$\frac{8}{j} \geqslant \frac{\gamma}{2^{22}}$$

assuming the first alternative gives the maximum in j. Certainly, taking $d = 1/4$ is permitted for $p \geqslant 17$, and gives the value (1.1).

Step 2 (how large is "large enough"). We gather here, as a series of inequalities to be satisfied by p, the conditions under which we can apply the previous lower bound. These we gather from the proofs of the results of this section. First come inequalities that make explicit the condition that $|G|$ be large enough in Theorem 4.2, , which are easily translated into conditions on p since $|\operatorname{SL}_2(\mathbf{F}_p)| = p(p^2-1)$.

- In order that (4.11) contradict (4.2), we must have

$$|G|^{\gamma - \delta_1(1+c_2)} > 4c_1.$$

- In order that (4.14) contradict the growth alternative of Helfgott's Theorem, it is enough that

$$|G|^{\gamma_1} > 4c_1 \Big\{ c_1 (4c_1)^{\gamma_1^{-1}} \Big\}^{(\delta - c_2\delta_1\gamma_1^{-1})^{-1}}$$

  where[9] $\gamma_1 = \gamma - (1+c_2)\delta_1$ (in view of (4.13)).
- In order that (4.16) give (4.3) when δ_1 satisfies (4.4), it is enough that

$$|G|^{\varepsilon - 2c_2\delta_1} \geqslant (\log 3|G|)^4,$$

  and that

$$|G|^{\delta_1} \geqslant c_1^{-2} (\log 3|G|)^4.$$

- In order that (4.17) hold, we must have

$$\min(|G|^{d/8}, |G|^{\delta_1/2}) \geqslant 2c_3. \tag{4.30}$$

Now we list the conditions needed to apply the Bourgain-Gamburd criterion in the situation of Theorem 1.1, when S freely generates a free group of rank $|S|/2 \geqslant 2$.

- We need

$$p \geqslant \max\Big(17, 2\exp\Big(\frac{c\tau(\log|S|)^2}{8}\Big)\Big),$$

  by (4.22).
- In order that the last inequality in (4.28) hold, as well as (4.29), it is enough that

$$|\operatorname{SL}_2(\mathbf{F}_p)|^{\gamma} \geqslant \max\Big(120, \Big(\log\frac{p}{2}\Big)\Big).$$

Remark 4.9. Below in Section 4.6 is found a straightforward Pari/GP [15] that computes the lower-bound of Step 1 for the spectral gap, given the set of matrices S, and that can also be used to determine for which p the bound is known to be applicable.

[9] This is not the same γ_1 that occurs in the proof of the decay of probabilities.

We finally explain to reduce the full statement of Theorem 1.1 to the case where the given symmetric subset $S \subset \mathrm{SL}_2(\mathbf{Z})$ generates a free group, which is the one treated by the Bourgain-Gamburd method.

For a given $S \subset \mathrm{SL}_2(\mathbf{Z})$ which generates a Zariski-dense subgroup G of SL_2, the intersection $G\cap\Gamma(2)$, where $\Gamma(2)$ is the principal congruence subgroup modulo 2, is a free subgroup of finite index in G. From a free generating set, one can extract two generators s_1, $s_2 \in G$ to obtain a free subgroup of rank 2 of G, say G_1 (since $G \cap \Gamma(2)$ has finite index in G, it is still Zariski-dense, and hence has rank at least 2). This subgroup is still Zariski-dense. We can then compare the expansion for the Cayley graphs of $\mathrm{SL}_2(\mathbf{F}_p)$ with respect to S and to $S_1 = \{s_1^{\pm 1}, s_2^{\pm 1}\}$.

For p large enough so that $G_p = \mathrm{SL}_2(\mathbf{F}_p)$ is generated both by S modulo p and S_1 modulo p, we have

$$d(x, y) \leqslant C d_1(x, y)$$

where $d_1(\cdot,\cdot)$ is the distance in the Cayley graph $\Gamma_1 = \mathcal{C}(G_p, S_1 \,(\mathrm{mod}\, p))$, and $d(\cdot,\cdot)$ the distance in $\Gamma_2 = \mathcal{C}(G_p, S \,(\mathrm{mod}\, p))$ and C is the maximum of the word length of s_1, s_2 with respect to S. Hence, by a standard lemma (see, e.g., [11, Lemma 3.1.16], applied to Γ_1 and Γ_2 with f the identity), the expansion constants satisfy

$$h(\mathcal{C}(G_p, S \,(\mathrm{mod}\, p))) = h(\Gamma_2) \geqslant w^{-1} h(\mathcal{C}(G_p, S_1 \,(\mathrm{mod}\, p)))$$

with

$$w = 4 \sum_{j=1}^{\lfloor C \rfloor} |S|^{j-1}.$$

In particular, using Theorem 1.1 for G_1, we obtain the expansion property for G, and we can bound the spectral explicitly once we know expressions for the generators s_1, s_2 in terms of those in S.

4.4. **Diameter bound.** We can now also prove quickly Corollary 1.4. Let $S_1 = S \cup \{1\}$. By Proposition 4.5, if we let

$$r = \left\lceil \tau \log \frac{p}{2} \right\rceil,$$

where τ is defined by (4.19), the size of ${S_1}^{(r)}$ is at least the size of a ball of radius r in a $|S|$-regular tree, which is well-known to be at least s^r, where $s = |S| - 1$.

For $p \geqslant 17$, this gives

$${S_1}^{(r)} \geqslant s^{-1} \left(\frac{p}{2}\right)^r \geqslant s^{-1} |\,\mathrm{SL}_2(\mathbf{F}_p)|^{\tau(\log s)/4},$$

and if $p \geqslant \exp(2\tau^{-1})$, this becomes

$${S_1}^{(r)} \geqslant |\,\mathrm{SL}_2(\mathbf{F}_p)|^{\delta_2},$$

where

$$\delta_2 = \frac{\tau(\log s)}{8} > 0.$$

Now we apply repeatedly Helfgott's Theorem with $H = {S_1}^{(r)}$. For j such that

$$j \geqslant \frac{\log(\delta_2^{-1})}{\log(1+\delta)},$$

the 3^j-fold product of H must be equal to $\mathrm{SL}_2(\mathbf{F}_p)$, and hence we get

$$\operatorname{diam} \mathcal{C}(\mathrm{SL}_2(\mathbf{F}_p), S) \leqslant 3^j r \leqslant 3^{j-1} (\log |\,\mathrm{SL}_2(\mathbf{F}_p)|),$$

and taking

$$j = \left\lceil \frac{\log(\delta_2^{-1})}{\log(1+\delta)} \right\rceil,$$

this gives the bound

$$\operatorname{diam} \mathcal{C}(\mathrm{SL}_2(\mathbf{F}_p), S) \leqslant 3^{\log(\delta_2^{-1})/\log(1+\delta)} (\log |\mathrm{SL}_2(\mathbf{F}_p)|).$$

4.5. **The Lubotzky group.** We can now prove Corollary 1.6. The group L generated by

$$S = \left\{ \begin{pmatrix} 1 & \pm 3 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ \pm 3 & 1 \end{pmatrix} \right\} \subset \mathrm{SL}_2(\mathbf{Z})$$

is a free group of rank 2. For all $p \neq 3$, the reduction of S modulo p generates $\mathrm{SL}_2(\mathbf{F}_p)$.

We have $\tau^{-1} = \log \sqrt{11}$ in that case, and $r = 2/\sqrt{3}$, so that (arguing directly with r) we compute that we have

$$\gamma_1 = \frac{\log(2/\sqrt{3})}{8 \log \sqrt{11}} = 0.014996592\ldots, \quad \text{for} \quad p \geqslant 17.$$

We further obtain

$$k \approx 0.026 \log(p/2), \qquad \gamma = 0.001129\ldots,$$

and the lower bound

$$\lambda_1(\Gamma_p) \geqslant 2^{-32}$$

for $\Gamma_p = \mathcal{C}(\mathrm{SL}_2(\mathbf{F}_p), S_p)$ and all p large enough.

Now using the script in the next section to determine the conditions that amount to p being large enough, we find that (4.30) is by far the worse condition, and holds for

$$p \geqslant 2^{2^{47}}.$$

We then find that the exponent A in Corollary 1.4 is bounded by 1563, which gives the diameter bound for the Lubotzky group.

4.6. **Script.** Here is a Pari/GP [15] script that performs the computations needed to obtain an explicit spectral for Theorem 1.1, given as input a set of matrices S which generate a free group (this condition is not checked).

```
\\ Norm of a matrix
matnorm(m)=sqrt(sum(i=1,matsize(m)[1],sum(j=1,matsize(m)[2],m[i,j]^2)))

\\ Spectral radius of random walk on k-regular tree
gapr(s)=local(k);k=length(s);k/2/sqrt(k-1)

\\ Growth constant in Helfgott's Theorem
gapdelta(s)=1/1344

\\ Minimal dimension of irreducible, OK for p at least 17
gapd(s)=1/4

\\ Constant c_2 in explicit multiplicative combinatorics
gapc2(s)=937

\\ Logarithm of c_1, base 2
gaplogc1(s)=2112

\\ Logarithm of c_3, base 2
```

```
gaplogc3(s)=2+2*gaplogc1(s)

\\ "tau" invariant
gaptau(s)=1/log(vecmax(vector(length(s),i,matnorm(s[i]))))

\\ How far to go to obtain decay of size gamma
gapc(s)=gaptau(s)/96

\\ Value of gamma for p large enough
gapgamma(s)=gaptau(s)*log(gapr(s))/8/32

\\ Value of gamma1 for p large enough
gapgamma1(s)=gapgamma(s)-gapdelta1(s)*(1+gapc2(s))

\\ Value of delta_1
gapdelta1(s)=1/2*min(gapdelta(s)*gapgamma(s)/(2*gapc2(s)+1),gapd(s)/8/
    gapc2(s))

\\ How many times one needs to iterate the Bourgain-Gamburd inequality
gapj(s)=8*max((gapc2(s)+1)/gapdelta(s)/gapgamma(s),8*gapc2(s)/gapd(s))

\\ Logarithm of the spectral gap for p large enough
gaploggap(s)=-gapd(s)/4/gapj(s)/gapc(s)

\\ First minimal value on log p, base 2
gaplogmin1(s)=(2+gaplogc1(s))/3/(gapgamma(s)-gapdelta1(s)*(1+gapc2(s)))

\\ Second minimal value on log p, base 2
gaplogmin2(s)=1/3/gapgamma1(s)*(2+gaplogc1(s)+1/(gapdelta(s)-gapc2(s)*
    gapdelta1(s)/gapgamma1(s))*(gaplogc1(s)+1/gapgamma1(s)*(2+gaplogc1(s
    ))))

\\ Is log(p) larger than third minimal value on log p (base e)?
gapislogmin3(s,lp)=if(lp>=1/3/(gapd(s)/4-2*gapc2(s)*gapdelta1(s))*4*(
    log(log(3)+lp)),1,0)

\\ Is log(p) larger than fourth minimal value on log p (base e)?
gapislogmin4(s,lp)=if(lp>=1/3/gapdelta1(s)*(4*log(log(3)+lp)-2*log(2)*
    gaplogc1(s)),1,0)

\\ Fifth minimal values on log p, base 2
gaplogmin5(s)=gaplogc3(s)/min(gapd(s)/8,gapdelta1(s)/2)

\\ Sixth minimal value of p
gaplogmin6(s)=log(max(17,2*gapr(s)^(2/gaptau(s)/gapr(s))))/log(2)

\\ Is log(p) larger than seventh minimal value on log p, base e
gapislogmin7(s,lp)=if(3*lp*gapgamma(s)>=log(lp-log(2)),1,0)

\\ Eighth minimal value on log p, base 2
gaplogmin8(s)=log(120)/log(2)/3/gapgamma(s)

\\ Exponent for diameter bound
```

```
gapdiamexp(s)=log(8/gaptau(s)/(length(s)-1))/log(1+gapdelta(s))

\\ Minimum of log(p), base 2, for gapislogmin3
gapfind3(s)= {
  local(j=2,i,k);
  while(!gapislogmin3(s,j),j=2*j);
  k=j/2;
  i=ceil((j+k)/2);
  while(i!=j,
    if(!gapislogmin3(s,i),
       k=i;i=ceil((j+k)/2),
       j=i;i=ceil((j+k)/2)));
  ceil(i/log(2));
}

\\ Minimum of log(p), base 2, for gapislogmin4
gapfind4(s)= {
  local(j=2,i,k);
  while(!gapislogmin4(s,j),j=2*j);
  k=j/2;
  i=ceil((j+k)/2);
  while(i!=j,
    if(!gapislogmin4(s,i),
       k=i;i=ceil((j+k)/2),
       j=i;i=ceil((j+k)/2)));
  ceil(i/log(2));
}

\\ Minimum of log(p), base 2, for gapislogmin7
gapfind7(s)= {
  local(j=2,i,k);
  while(!gapislogmin7(s,j),j=2*j);
  k=j/2;
  i=ceil((j+k)/2);
  while(i!=j,
    if(!gapislogmin7(s,i),
       k=i;i=ceil((j+k)/2),
       j=i;i=ceil((j+k)/2)));
  ceil(i/log(2));
}

\\ Minimum value of log(p), base 2
gapmin(s)=ceil(vecmax([gaplogmin1(s),gaplogmin2(s),gapfind3(s),gapfind4
   (s),gaplogmin5(s),gaplogmin6(s),gapfind7(s),gaplogmin8(s)]))

\\ Base 2 bound for gapmin(s)
gapminlog(s)=ceil(log(gapmin(s))/log(2))


\\ Generators of the Lubotzky group
ls=[[1,3;0,1],[1,-3;0,1],[1,0;3,1],[1,0;-3,1]]

\\ ? gaploggap(ls)
```

```
\\ %30 = -3.342735349226749447346680705300590888 2 E-10
\\ ? log(-gaploggap(ls))/log(2)
\\ %31 = -31.47825181033232651674591639494092723 3
\\ ? gapminlog(ls)
\\ %32 = 47
```

5. Appendix: proof of Theorem 2.1

In this appendix, we sketch the proof of Theorem 2.1, following very (essentially) line by line Tao's paper [18]. The presentation is therefore highly condensed, though we use a "diagram" notation which should make it relatively easy to check how the values of the constants evolve.

Below all sets are subsets of a fixed finite group G, and are all non-empty.

5.1. **Diagrams.** We will use the following diagrammatic notation:

(1) If A and B are sets with Ruzsa distance

$$d(A,B) = \log\Big(\frac{|A||B^{-1}|}{\sqrt{|A||B|}}\Big)$$

such that $d(A,B) \leqslant \log\alpha$, we write

$$A \underset{\alpha}{\bullet\!\!-\!\!-\!\!-\!\!\bullet} B\ ,$$

(2) If A and B are sets with $|B| \leqslant \alpha|A|$, we write

$$B \overset{\alpha}{\bullet\!\!\longrightarrow} A\ ,$$

and in particular if $|X| \leqslant \alpha$, we write $X \overset{\alpha}{\bullet\!\!\longrightarrow} 1$,

(3) If A and B are sets with $e(A,B) \geqslant 1/\alpha$, we write

$$A \overset{\alpha}{\bullet\!\!\leadsto\!\!\bullet} B\ ,$$

(4) If $A \subset B$, we also write $A \rightarrowtail B$.

The following rules are easy to check (in addition to some more obvious ones which we do not spell out):

(1) From

$$A \underset{\alpha}{\bullet\!\!-\!\!-\!\!-\!\!\bullet} B$$

we can get

$$A \overset{\alpha^2}{\bullet\!\!\longrightarrow} B\ , \qquad B \overset{\alpha^2}{\bullet\!\!\longrightarrow} A\ .$$

(2) (Ruzsa's triangle inequality) From

$$A \underset{\alpha_1}{\bullet\!\!-\!\!-\!\!-\!\!\bullet} B \underset{\alpha_2}{\bullet\!\!-\!\!-\!\!-\!\!\bullet} C$$

we get

$$A \underset{\alpha_1\alpha_2}{\bullet\!\!-\!\!-\!\!-\!\!\bullet} C\ .$$

(3) From

$$C \overset{\alpha_1}{\bullet\!\!\longrightarrow} B \overset{\alpha_2}{\bullet\!\!\longrightarrow} A$$

we get

$$C \overset{\alpha_1\alpha_2}{\bullet\!\!\longrightarrow} A\ .$$

(4) ("Unfolding edges") From

$$B \overset{\alpha}{\bullet\!\!\longrightarrow} A \qquad B \underset{\beta}{\bullet\!\smile\!\bullet} A$$

we get

$$AB^{-1} \overset{\sqrt{\alpha}\beta}{\bullet\!\!\longrightarrow} A$$

(note that by the second point in this list, we only need to have

$$B \underset{\beta}{\bullet\!\!-\!\!\bullet} A$$

to obtain the full statement with $\alpha = \beta^2$, which is usually qualitatively equivalent.)

(5) ("Folding") From

$$AB^{-1} \overset{\alpha}{\bullet\!\!\longrightarrow} A \overset{\beta}{\bullet\!\!\longrightarrow} B$$

we get

$$A \underset{\alpha\beta^{1/2}}{\bullet\!\!\longrightarrow\!\!\bullet} B \; .$$

Note that the relation $A \overset{\alpha}{\bullet\!\!\longrightarrow} B$ is purely a matter of the size of A and B, while the other arrow types depend on structural relations involving the sets (for $A \rightarrowtail B$) and product sets (for $A \underset{\alpha}{\bullet\!\!-\!\!\bullet} B$ or $A \overset{\alpha}{\bullet\!\!\leadsto\!\!\bullet} B$).

5.2. **Proofs.** First we state the Ruzsa covering lemma [18, Lemma 3.6] in our language:

Theorem 5.1 (Ruzsa)**.** *If*

$$AB \overset{\alpha}{\bullet\!\!\longrightarrow} A \; ,$$

there exists a set X which satisfies

$$X \rightarrowtail B \; , \qquad X \overset{\alpha}{\bullet\!\!\longrightarrow} 1 \; , \qquad B \rightarrowtail A^{-1}AX \; ,$$

and symmetrically, if

$$BA \overset{\alpha}{\bullet\!\!\longrightarrow} A \; ,$$

there exists Y with

$$Y \rightarrowtail B \; , \qquad Y \overset{\alpha}{\bullet\!\!\longrightarrow} 1 \; , \qquad B \rightarrowtail XAA^{-1} \; .$$

Next we have the link between sets with small tripling and approximate subgroups [18, Th. 3.9 and Cor. 3.10]:

Theorem 5.2. *Let $A = A^{-1}$ with $1 \in A$ and*

$$A^{(3)} \overset{\alpha}{\bullet\!\!\longrightarrow} A \; .$$

Then $H = A^{(3)}$ is a $(2\alpha^{44})$-approximate subgroup containing A.

Proof. We have first

$$H \overset{\alpha}{\bullet\!\!\longrightarrow} A \; , \qquad A \rightarrowtail H \; .$$

Then by Ruzsa's lemma 3.1, we get

$$AH^{(2)} = A^{(7)} \overset{\alpha^5}{\bullet\!\!\longrightarrow} A \; ,$$

and by the Ruzsa covering lemma there exists X with

$$X \rightarrowtail H^{(2)} \; , \qquad X \overset{\alpha^5}{\bullet\!\!\longrightarrow} 1 \; ,$$

such that

$$H^{(2)} \rightarrowtail A^{(2)}X \rightarrowtail A^{(3)}X = HX \cdot$$

Taking $X_1 = X \cup X^{-1}$, we get

$$X_1 \rightarrowtail H^{(2)}, \qquad X_1 \bullet\!\xrightarrow{2\alpha^5} 1,$$

and

$$H^{(2)} \rightarrowtail HX, \qquad H^{(2)} \rightarrowtail XH,$$

which are the properties defining a $(2\alpha^5)$-approximate subgroup. ☐

The next result is the explicit form of [18, Th. 4.6, (i) implies (ii)]:

Theorem 5.3. *Let A and B with*

$$A \underset{\alpha}{\bullet\!\!-\!\!\!-\!\!\bullet} B^{-1}$$

Then there exists a γ-approximate subgroup H and a set X with

$$X \bullet\!\xrightarrow{\gamma_1} 1, \qquad A \rightarrowtail XH, \qquad B \rightarrowtail HX, \qquad H \bullet\!\xrightarrow{\gamma_2} A,$$

where

$$\gamma \leqslant 2^{21}\alpha^{80}, \qquad \gamma_1 \leqslant 2^{28}\alpha^{104}, \qquad \gamma_2 \leqslant 8\alpha^{14}.$$

Furthermore, one can ensure that

$$H^{(3)} \bullet\!\xrightarrow{2^{10}\alpha^{40}} H \cdot \tag{5.1}$$

Proof. From

$$A \underset{\alpha^2}{\overset{1}{\bullet\!\!\longrightarrow}} A,$$

we get first

$$AA^{-1} \bullet\!\xrightarrow{\alpha^2} A \cdot$$

By [18, Prop. 4.5], we find a set S with[10] $1 \in S$ and $S = S^{-1}$ such that

$$A \bullet\!\xrightarrow{2\alpha^2} S, \qquad AS^{(n)}A^{-1} \bullet\!\xrightarrow{2^n\alpha^{4n+2}} A$$

for all $n \geqslant 1$. In particular, we get

$$AS^{-1} = AS \bullet\!\xrightarrow{2\alpha^6} A, \qquad S \bullet\!\xrightarrow{2\alpha^6} A \cdot$$

Thus

$$AS^{-1} \bullet\!\xrightarrow{2\alpha^6} A \bullet\!\xrightarrow{2\alpha^6} S,$$

which gives

$$A \underset{\beta}{\bullet\!\!-\!\!\!-\!\!\bullet} S$$

by folding, with $\beta = 2\sqrt{2}\alpha^7$.

In addition, we have

$$S^{(3)} \bullet\!\xrightarrow{8\alpha^{14}} A \bullet\!\xrightarrow{2\alpha^2} S,$$

[10] The property $1 \in S$ is not explicitly stated in [18], but follows from the explicit definition used by Tao, namely $S = \{x \in G \mid |A \cap Ax| > (2\alpha^2)^{-1}|A|\}$.

and Theorem 5.2 says that $H = S^{(3)}$ is a γ-approximate subgroup containing S, with $\gamma = 2(16\alpha^{16})^5 = 2^{21}\alpha^{80}$, and (as we see)

$$H \bullet\!\xrightarrow{\;8\alpha^{14}\;} A \ .$$

Moreover, we have

$$H^{(3)} = S^{(9)} \rightarrowtail AS^{(9)}A^{-1} \bullet\!\xrightarrow{\;2^9\alpha^{38}\;} A \bullet\!\xrightarrow{\;2\alpha^2\;} S \ ,$$

which gives (5.1).

Now from

$$AH = AS^{(3)} \bullet\!\xrightarrow{\;8\alpha^{14}\;} A \bullet\!\xrightarrow{\;2\alpha^2\;} S \bullet\!\xrightarrow{\;1\;} H \ ,$$

we see by the Ruzsa covering lemma that there exists Y with

$$Y \rightarrowtail A \ , \quad Y \bullet\!\xrightarrow{\;16\alpha^{16}\;} 1 \ , \quad A \rightarrowtail YHH \ .$$

By definition of an approximate subgroup, there exists Z with

$$Z \bullet\!\xrightarrow{\;\gamma\;} 1 \ , \quad HH \rightarrowtail ZH \ ,$$

and hence

$$A \rightarrowtail (YZ)H \ .$$

Now we go towards B. First we have

$$AH^{-1} = AS^{(3)} \bullet\!\xrightarrow{\;8\alpha^{14}\;} A \bullet\!\xrightarrow{\;\alpha^2\;} H$$

which, again by folding, gives

$$A \bullet\!\underset{\alpha_1}{\text{---}}\!\bullet H$$

with $\alpha_1 = 8\sqrt{2}\alpha^{15}$. Hence we can write

$$H \bullet\!\underset{\alpha_1}{\text{---}}\!\bullet A \bullet\!\underset{\alpha}{\text{---}}\!\bullet B^{-1} \ ,$$

and so

$$H \bullet\!\underset{\alpha\alpha_1}{\text{------}}\!\bullet B^{-1} \ .$$

In addition, we have

$$H \bullet\!\xrightarrow{\;8\alpha^{14}\;} A \bullet\!\xrightarrow{\;\alpha^2\;} B^{-1} \ ,$$

and therefore we get

$$H \bullet\!\xrightarrow{\;8\alpha^{16}\;} B^{-1}, \qquad H \bullet\!\underset{\alpha\alpha_1}{\text{------}}\!\bullet B^{-1}$$

from which it follows by unfolding that

$$B^{-1}H^{-1} = B^{-1}H \bullet\!\xrightarrow{\;32\alpha^{20}\;} B^{-1} \bullet\!\xrightarrow{\;\alpha^2\;} A \bullet\!\xrightarrow{\;2\alpha^2\;} H \ .$$

Once more by the Ruzsa covering lemma, we find Y_1 with

$$Y_1 \rightarrowtail B^{-1} \ , \quad Y_1 \bullet\!\xrightarrow{\;32\alpha^{20}\;} 1 \ , \quad B^{-1} \rightarrowtail Y_1HH \rightarrowtail (Y_1Z)H \ .$$

Now we need only take $X = (Y_1Z \cup YZ)$, so that

$$X \bullet\!\xrightarrow{\;\gamma_1\;} 1$$

with $\gamma_1 = \gamma(64\alpha^{24} + 16\alpha^{16})$, in order to conclude. Since

$$\gamma_1 \leqslant 2^{28}\alpha^{104},$$

we are done. $\square$

The next result is a version of the Balog-Gowers-Szemerédi Lemma found in [18, Th. 5.2].

Theorem 5.4. *Let A and B with*

$$A \bullet\!\overset{\alpha}{\rightsquigarrow}\!\bullet B\ .$$

Then there exist A_1, B_1 with

$$A_1 \rightarrowtail A\ ,\qquad B_1 \rightarrowtail B\ ,$$

as well as

$$A \bullet\!\xrightarrow{\ 8\sqrt{2}\alpha\ } A_1\ ,\qquad B \bullet\!\xrightarrow{\ 8\alpha\ } B_1\ ,$$

and

$$A_1 \bullet\!\underset{\alpha_1}{\text{———}}\!\bullet B_1^{-1}$$

where $\alpha_1 = 2^{20}\alpha^9$.

This is not entirely spelled out in [18], but only the last two or three inequalities in the proof need to be made explicit to obtain this value of α_1. Finally, the next theorem is just the "diagrammatic" version of Theorem 2.1, and therefore completes its proof. It is an explicit version of [18, Th. 5.4; (i) implies (iv)].

Theorem 5.5. *Let A and B with*

$$A \bullet\!\overset{\alpha}{\rightsquigarrow}\!\bullet B\ .$$

Then there exist a β-approximate subgroup H and x, $y \in G$, such that

$$H \bullet\!\xrightarrow{\beta_2} A\ ,\qquad A \bullet\!\xrightarrow{\beta_1} A \cap xH\ ,\qquad B \bullet\!\xrightarrow{\beta_1} B \cap Hy\ ,$$

where

$$\beta \leqslant 2^{1621}\alpha^{720},\qquad \beta_1 \leqslant 2^{2112}\alpha^{937},\qquad \beta_2 \leqslant 2^{283}\alpha^{126}.$$

Moreover, one can ensure that

$$H^{(3)} \bullet\!\xrightarrow{\beta_3} H$$

where $\beta_3 = 2^{810}\alpha^{360}$.

Proof. By the Balog-Gowers-Szemerédi Theorem, we get A_1, B_1 with

$$A_1 \rightarrowtail A\ ,\qquad B_1 \rightarrowtail B\ ,$$

as well as

$$A \bullet\!\xrightarrow{\ 8\sqrt{2}\alpha\ } A_1\ ,\qquad B \bullet\!\xrightarrow{\ 8\alpha\ } B_1\ ,$$

and

$$A_1 \bullet\!\underset{\alpha_1}{\text{———}}\!\bullet B_1^{-1}$$

where $\alpha_1 = 2^{20}\alpha^9$. Applying Theorem 5.3 to A_1 and B_1, we get a β-approximate subgroup H and a set X with

$$H \bullet\!\xrightarrow{\ 2\alpha_1^{14}\ } A_1 \bullet\!\xrightarrow{1} A$$

and

$$X \bullet\!\xrightarrow{\gamma} 1\ ,\qquad A_1 \rightarrowtail XH\ ,\qquad B_1 \rightarrowtail HX\ ,$$

where

$$\beta = 2^{21}\alpha_1^{80} = 2^{1621}\alpha^{720},\qquad \gamma = 2^{28}\alpha_1^{104} = 2^{2108}\alpha^{936},$$

and moreover

$$H^{(3)} \overset{\beta_3}{\bullet\!\!\longrightarrow} H$$

where $\beta_3 = 2^{10}\alpha_1^{40} = 2^{810}\alpha^{360}$.

Applying the pigeonhole principle, we find x such that

$$A \overset{8\sqrt{2}\alpha}{\bullet\!\!\longrightarrow} A_1 \overset{\gamma}{\bullet\!\!\longrightarrow} A_1 \cap xH \rightarrowtail A \cap xH$$

and y with

$$B \overset{8\alpha}{\bullet\!\!\longrightarrow} B_1 \overset{\gamma}{\bullet\!\!\longrightarrow} B_1 \cap Hy \rightarrowtail B \cap Hy\ .$$

This gives what we want with

$$\beta_1 \leqslant 8\sqrt{2}\alpha\gamma \leqslant 2^{2112}\alpha^{937}, \qquad \beta_2 = 8\alpha_1^{14} = 2^{283}\alpha^{126}.$$

□

References

[1] J. Bourgain and A. Gamburd: *Uniform expansion bounds for Cayley graphs of* $\mathrm{SL}_2(\mathbf{F}_p)$, Ann. of Math. 167 (2008), 625–642.

[2] J. Bourgain, A. Gamburd and P. Sarnak: *The affine linear sieve*, Invent. math. 179 (2010), 559–644.

[3] E. Breuillard, B. Green and T. Tao: *Approximate subgroups of linear groups*, GAFA 21 (2011), 774–819; `arXiv:1005.1881`.

[4] A. Gamburd: *On the spectral gap for infinite index "congruence" subgroups of* $\mathrm{SL}_2(\mathbf{Z})$, Israel J. Math. 127 (2002), 157-–200.

[5] A.A. Glibichuk and S.V. Konyagin: *Additive properties of product sets in fields of prime order*, in "Additive Combinatorics", C.R.M. Proc. and Lecture Notes 43, A.M.S (2006), 279–286.

[6] W.T. Gowers: *Quasirandom groups*, Comb. Probab. Comp. 17 (2008), 363–387.

[7] H. Helfgott: *Growth and generation in* $\mathrm{SL}_2(\mathbf{Z}/p\mathbf{Z})$, Ann. of Math. 167 (2008), 601–623.

[8] E. Hrushovski: *Stable group theory and approximate subgroups*, Journal of the A.M.S 25 (2012), 189–243.

[9] M. Huxley: *Exceptional eigenvalues and congruence subgroups*, in "The Selberg trace formula and related topics", p. 341– 349; edited by D. Hejhal, P. Sarnak and A. Terras, Contemporary Math. 53, A.M.S, 1986.

[10] E. Kowalski: *Sieve in expansion*, Séminaire, exposé 1028 (2010); to appear in Astérisque.

[11] E. Kowalski: *Expander graphs*, lecture notes for ETH Zürich Fall Semester course; `www.math.ethz.ch/~kowalski/expander-graphs.pdf`

[12] M. Larsen and R. Pink: *Finite subgroups of algebraic groups*, Journal of the A.M.S 24 (2011), 1105–1158.

[13] D. Levin, Y. Peres and E. Wilmer: *Markov chains and mixing times*, A.M.S 2009.

[14] N. Nikolov and L. Pyber: *Product decompositions of quasirandom groups and a Jordan-type theorem*, J. European Math. Soc. 13 (2011), 1063–1077.

[15] PARI/GP, version `2.6.0`, Bordeaux, 2011, `http://pari.math.u-bordeaux.fr/`.

[16] L. Pyber and E. Szabó: *Growth in finite simple groups of Lie type of bounded rank*, preprint (2010), `arXiv:1005.1858v1`

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

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

[19] T. Tao and V. Vu: *Additive combinatorics*, Cambridge Studies Adv. Math. 105, Cambridge Univ. Press (2006).

[20] P. Varjú: *Expansion is* $SL_d(O_K/I)$, preprint (2010).

ETH Zürich – D-MATH, Rämistrasse 101, 8092 Zürich, Switzerland

E-mail address: `kowalski@math.ethz.ch`