Geom. Funct. Anal. Vol. 23 (2013) 1–41
DOI: 10.1007/s00039-012-0200-9
Published online February 8, 2013

GAFA Geometric And Functional Analysis

EXPANSION IN $\mathrm{SL}_2(\mathbb{R})$ AND MONOTONE EXPANDERS

Jean Bourgain And Amir Yehudayoff

Abstract. This work presents an explicit construction of a family of monotone expanders, which are bi-partite expander graphs whose edge-set is defined by (partial) monotone functions. The family is (roughly) defined by the Möbius action of $\mathsf{SL}_2(\mathbb{R})$ on the interval [0,1]. A key part of the proof is a product-growth theorem for certain subsets of $\mathsf{SL}_2(\mathbb{R})$. This extends recent results on finite/compact groups to the non-compact scenario. No other proof-of-existence for monotone expanders is known.

1 Introduction

1.1 Expander graphs. Expanders are sparse graphs with "strong connectivity" properties. Such graphs are extremely useful and are basic tools, for example, in constructions of error correcting codes [SS96] and of compressed sensing matrices [GLW08]. For the vast applications of expander graphs, see the survey [HLW06]. Most sparse graphs are expanders, but for applications explicit (simple to describe) constructions of expanders are required. Indeed, explicit constructions of expanders graphs are known, both using algebraic methods (e.g. [LPS88]) and using combinatorial methods (e.g. [RVW02]).

In a nutshell, explicit expanders are so useful because of their two contradictory faces: on one hand, expander graphs behave, in many ways, like random graphs, and random objects are, for example, typically hard to construct. On the other hand, the word explicit means that they in fact can be easily described and constructed.

A natural question that is potentially of practical importance is thus "how simple can expander graphs be?" We now discuss two properties of the expanders we construct that show that there are "simple" expanders. One of the first explicit constructions of expander graphs was given by Gabber and Galil [GG79]. The Gabber–Galil expander is extremely simple: It has a two-dimensional torus as a vertex-set and its edges are defined by simple linear functions on the torus. Klawe, on the other hand, showed that no one-dimensional analog of the Gabber–Galil expander

Horev fellow—supported by the Taub foundation. This research was supported by Sindey & Ann Grazi Research Fund. Research partially supported by ISF and BSF.

exists [Kla84]. The expanders we construct are "close to being" one-dimensional (formal definitions follow). Another well-known fact is that planar graphs can not be expanders [Ung51]. From a different perspective, a monotone expander is "close" to being planar: it is a finite union of planar graphs, all of which respect the same planar embedding of vertices and their edges can be drawn as straight line segments.

As mentioned, in many cases as well as in the case of expanders, proving the existence of an object is much easier than constructing it. In the case of monotone expanders, however, no non-explicit proof of existence is known, and the only proof of existence known is the explicit construction presented here (Dvir and Wigderson [DW10] showed, nevertheless, that any proof of existence of a family of monotone expanders yields an explicit construction of monotone expanders). A partial explanation to that is the following. Natural probability distributions on (partial) monotone functions give, w.h.p., functions that are "close" to affine. Klawe, however, showed in [Kla84] that if one tries to construct expanders using affine transformations, then the minimal number of generators required is super-constant, and so no construction "that is close to affine" can work. The construction in this text uses edges that are defined as the ratio of two affine transformations, and so the edges are slightly more elaborate than what is impossible by Klawe's result.

1.2 Monotone expander graphs. The construction of monotone expanders we present first builds a "continuous monotone" expander, which in turn can be discretized to the required size. A *continuous* expander is a finite family of maps Ψ for which there exists a constant $c > 0$ so that the following holds. Every $\psi \in \Psi$ is a smooth map from a sub-interval of $[0,1]$ to a sub-interval of $[0,1]$, and for all measurable $A \subset [0,1]$ with $|A| \leq 1/2$,

$$|\Psi(A)| \geq (1+c)|A|,$$

where $\Psi(A) = \bigcup_{\psi \in \Psi} \psi(A)$. One way to think of a continuous expander is as an infinite constant degree bi-partite graph with two color-classes that are copies of $[0,1]$, where Ψ defines the edges between the color-classes. We say that Ψ is *monotone* if in addition every $\psi \in \Psi$ is monotone, namely, $\psi(x) > \psi(y)$ for $x > y$. Pictorially, this means that if drawn on the plane with the two color-classes as two parallel straight line segments of length one, and with edges drawn as straight line segments as well, then for every ψ in Ψ, the edges defined by ψ do not cross each other.

Theorem 1. *There exists an explicit continuous monotone expander.*

The word explicit in the theorem can be interpreted as follows. The family Ψ can be (uniformly) described by a constant number of bits, and given a rational $x \in [0,1]$ that can be described by b bits, $\psi(x)$ is rational and can be computed in time polynomial in b, for all $\psi \in \Psi$.

The theorem above describes the existence of a continuous monotone expander. By partitioning $[0,1]$ to n equal-length intervals, Ψ naturally defines a family of discrete monotone expanders. Namely, for every n, a graph $G = G_n$ satisfying the

following properties. First, G is bi-partite: the vertex-set of G is partitioned to two ordered sets L and R, each of size n. Second, G is (finite-degree) monotone: there exist an integer k independent of n, a family of sets $L_1, \ldots, L_k \subset L$, and a family of monotone maps $\{f_i : L_i \to R : 1 \leq i \leq k\}$ so that the edges of G are of the form $\{a, f_i(a)\}$ for $i \in [k]$ and $a \in L_i$. Finally, G is an expander: there exists a constant $c > 0$ independent of n so that for every $A \subset L$ of size $|A| \leq n/2$, the size of $B = \{b \in R : \{a, b\} \in E(G) \text{ for some } a \in A\}$, the neighborhood of A in R, is at least $(1 + c)|A|$.

COROLLARY 2. *For every integer n, there exists an explicit discrete bi-partite monotone expander on $2n$ vertices.*

Why does the theorem imply the corollary? We shall implicitly define the partial monotone maps $\{f_i\}$ by defining the edges in the discrete monotone expander graph. Given an integer n, set $L = R = [n]$. Partition $[0, 1]$ to n equal-length consecutive intervals $I_1, \ldots, I_n$. For every ψ in the continuous monotone expander Ψ, let J_ψ be the interval in $[0, 1]$ on which ψ is defined. Two elements $a \in L$ and $b \in R$ are connected by an edge iff $\psi(I_a \cap J_\psi) \cap I_b \neq \emptyset$ for some ψ in Ψ. Since all maps ψ are smooth, the length of the interval $\psi(I_a \cap L_\psi)$ is at most a constant times $1/n$. This implies that for every ψ in Ψ, the edges thus defined by ψ can be "covered" by a constant number of partial monotone maps from L to R. The total number of maps defining the discrete expander is therefore constant. It remains to prove expansion. Let A be a subset of L of size at most $n/2$ and let B be the neighborhood of A in R. Let A' be the subset of $[0, 1]$ that corresponds to A, i.e., $A' = \bigcup_{a \in A} I_a$. Let $B' = \bigcup_{b \in B} I_b$. By construction, $\Psi(A') \subseteq B'$. Since Ψ is a continuous monotone expander,

$$(1 + c)|A|/n = (1 + c)|A'| \leq |\Psi(A')| \leq |B'| = |B|/n.$$

1.3 Growth. The construction of monotone expanders using a matrix group fits well into recent developments on growth and expansion in matrix groups. The three steps of the proof correspond to the three steps of the proof in the work of Bourgain and Gamburd [BG07] showing expansion in $\mathsf{SU}(2)$. The proofs in [BG07] use ideas from the work of Bourgain and Gamburd [BG08] proving expansion in $\mathsf{SL}_2(\mathbb{F}_p)$, and the work of Helfgott [H08] showing growth in $\mathsf{SL}_2(\mathbb{F}_p)$.

Helfgott's work contains the first product-growth theorem for matrix groups. A key ingredient of the proofs in [BG07, BG08] is a product-growth theorem for finite or compact groups. We, too, prove a product-growth theorem, but for the non-compact $\mathsf{SL}_2(\mathbb{R})$. The proof idea is similar to previous works, but non-compactness introduces some technical and conceptual difficulties.

With the understanding that "typically" a Cayley graph of a matrix group is an expander, it is natural to try and use a matrix group to also define a monotone expander. Matrix groups, however, do not have a natural order on them. So, instead of a Cayley graph, use a graph defined by the group's action on some ordered set, like the real numbers. A well-known such action is the Möbius action of $\mathsf{SL}_2(\mathbb{R})$ on

the real numbers. As it turns out, this action is in fact (piece-wise) monotone as well. Indeed, this action will essentially define the monotone expander.

This line of thought for constructing monotone expanders was suggested in [Bou09], together with an outline of a proof. Here we provide a full proof.

1.4 Applications. *Dimension expanders.* Implicit in the work of Dvir and Shpilka [DS08] it is shown that an explicit discrete monotone expander easily yields an explicit family of dimension expanders over any field (see [DW10] as well). We mention that the work of Lubotzky and Zelmanov [LZ08] shows that over the real numbers many known families of (non-monotone) expander graphs similarly yield dimension expanders. The only known way to construct dimension expanders for general fields is using monotone expanders.

Turing machines. Another application of monotone expanders is proved in [DW10]. They showed that monotone expanders yield constant-page pushdown expanders, which are graphs that arise in certain Turing machine simulations.

Furstenberg measures. In [Bou12], it is shown that the finitely supported symmetric measure on the group $\mathsf{SL}_2(\mathbb{R})$ constructed here to prove Theorem 1 has a Furstenberg measure that is absolutely continuous with respect to the projective measure (and in fact with a density that can be made arbitrarily smooth). This questions is motivated by a conjecture due to Kaimanovich and Le Prince [KL10] and related works of Simon, Solomyak and Urbanski [SSU01a, SSU01b].

Schrödinger operators. The product-growth theorem for $\mathsf{SL}_2(\mathbb{R})$ we prove is exploited in [Bou12] to study the co-cycles of one-dimensional Schrödinger operators with a Hölder potential to prove the smoothness of the density of states in this context.

2 Proof Outline

We now describe the outline of the proof. We ignore many of the problematic and technical issues, and just present the flow of ideas (much work is required to transform this sketch into a full proof).

Notation. For convenience, we use the following notation throughout the text. For a constant $c \in \mathbb{R}$, we denote by $c+$ a constant slightly larger than c, and by $c-$ a constant slightly smaller than c. Typically, the meaning of "slightly" depends on other parameters that are clear from the context. We also use the following asymptotic notation. Write $a \lesssim b$ if $a \leq Cb$ with C a universal constant. Write $a \gtrsim b$ if $b \lesssim a$, and $a \sim b$ if $a \lesssim b \lesssim a$.

2.1 Defining maps. Consider the special linear group $\mathsf{SL}_2(\mathbb{R})$, the set of all 2×2 matrices with entries in $\mathbb{R}$ and determinant one. Every matrix $g \in \mathsf{SL}_2(\mathbb{R})$ acts on $\mathbb{R}$ in a monotone way via the Möbius action: if $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ in $\mathsf{SL}_2(\mathbb{R})$, then the action of g on x in the projective real line is

$$\overline{g}(x) = \frac{ax+b}{cx+d}.$$

The action is monotone as the derivative of $\overline{g}$ is positive, except at $-d/c$. For details, see Section 3.

The maps in Ψ will be (roughly) defined by the actions of a finite set of matrices $\mathcal{G} \subset \mathsf{SL}_2(\mathbb{R})$,

$$\Psi = \{\overline{g} : g \in \mathcal{G} \text{ or } g^{-1} \in \mathcal{G}\}.$$

This ensures that the maps in Ψ are monotone. It thus remains to choose $\mathcal{G}$. We, in fact, describe two choices of sets of generators $\mathcal{G}_0$ and $\mathcal{G}$ so that $\mathcal{G} \subset \mathcal{G}_0$ (this double choice will be explained in more detail later on). For this preliminary discussion, we focus on the set $\mathcal{G}$.

The set $\mathcal{G}$ will be chosen as a family of matrices that freely generate a group (with some extra properties, see Lemma 3 for exact statement). To find $\mathcal{G}$, use the strong Tits alternative of Breuillard [Bre08], which roughly states that every ball of constant radius in $\mathsf{SL}_2(\mathbb{Z})$ contains elements that freely generate a group.

2.2 Proving expansion: three steps. As in many expanders constructions and following recent works, the expansion is established by showing that the random walk/flow defined by Ψ is rapidly mixing: a sequence $g_1, g_2, \ldots$ of elements of $\mathcal{G}$ defines a walk $w_0, w_1, \ldots$ in $\mathsf{SL}_2(\mathbb{R})$ via $w_0 = 1$ and $w_{t+1} = g_t w_{t-1}$. It also defines a flow on $[0,1]$; A point x in $[0,1]$ flows to $x_0, x_1, \ldots$ that are defined by $x_0 = x$ and $x_t = \overline{g}_t(x_{t-1}) = \overline{w}_t(x)$. By rapidly mixing we (roughly) mean that for every x in $[0,1]$, if the sequence $g_1, g_2, \ldots$ is i.i.d. and uniform in $\mathcal{G}$, then w_t, x_t are close to being uniformly distributed[1] for relatively small t.

The proof of rapid mixing follows by showing that if at time t the walk/flow are not close to uniform, then at time $t+1$ they are closer to uniform than at time t. In other words, that the entropy strictly grows at each step. The entropy here is measured as the exponential of the Rényi entropy, that is, as $1/\|w_t\|_2^2$. In the proof, the reason for this increase in entropy changes with t. There are three different reasons for three different time phases: small t, intermediate-size t and large t. We now describe the idea of the proof for each of the three time phases.

(i) Small t. In the first phase we show that when t is small, w_{t+1} has more entropy than w_t. There are two ingredients to the proof of this statement: (a) the group generated by $\mathcal{G}$ is free, and (b) the "diophantine" property of $\mathcal{G}$, that is, elements of $\mathcal{G}$ have constant rational entries.

The freeness of the group generated by $\mathcal{G}$ tells us that the walk $w_1, w_2, \ldots$ corresponds to a random walk along a tree τ that is embedded in $\mathsf{SL}_2(\mathbb{R})$; The root of τ is 1 and a node v in τ is connected to all the possible values for gv for g in $\mathcal{G} \cup \mathcal{G}^{-1}$. Intuitively, since the entropy of a random walk along a tree increases at each step, the entropy of the flow increases at each step as well. To make the argument work, though, we need that by looking at the walk, we can recover the tree τ. This is

[1] Since $\mathsf{SL}_2(\mathbb{R})$ is not compact, there is no uniform distribution on it.

possible by property (b) that tells us that different nodes of τ are far away from each other. So, since t is small, we are able to recover the tree from watching the flow without "using a magnifying glass." This, in turn, shows that the entropy strictly grows at each step.

(ii) Intermediate-size t. This phase is the main part of the argument. As in previous works, this step follows via a product-growth theorem. We prove a product-growth theorem for $\mathsf{SL}_2(\mathbb{R})$: if S is a subset of $\mathsf{SL}_2(\mathbb{R})$ with certain properties, then the size of $S_{(3)} = \{s_1 s_2 s_3 : s_i \in S\}$ is much larger than the size of S. Below we explain some of the ideas from the proof of this statement, but we first explain how it is related to rapid mixing.

We argue that, as long as w_t is not close to uniform, w_{3t} has much more entropy than w_t. Think of S as the support of w_t. If w_t is not already close to uniform, then S will satisfy the conditions of the product-growth theorem, and so $S_{(3)}$ will be much larger than S. This, in turn, implies that w_{3t}, which corresponds to $S_{(3)}$, has much more entropy than w_t.

The proof of the product theorem consists of several parts (and its outline is similar to that of [BG07], but the proof in our case is more elaborate). Here we describe the flow of the argument (for the full proof, see Section 7). We wish to prove that a set S with certain properties becomes larger when multiplied by itself. The first step (which already appears in [H08]) is to use matrix-trace to move from matrix products to sums and products in $\mathbb{R}$; if g_1, g_2 are two matrices, then the trace of $g_1 g_2$ involves both sums and products in the field. Then, instead of arguing of matrix products, we can argue of field operations. To do so, we use a "sum-product" theorem for $\mathbb{R}$ called the discretized ring conjecture, which roughly states that a well-distributed set in $\mathbb{R}$ becomes larger under sums and products. So, if S satisfies certain properties, its trace-set will be well-distributed, and so the trace-set of $S_{(3)}$ will be much larger than that of S. This, in turn, is used to proved that $S_{(3)}$ is much larger than S.

(iii) Large t. By (i) and (ii), we can conclude that w_t has large entropy, for t relatively small. The final step of the proof follows by proving a "mixing property." In previous works, this last step follows Sarnak and Xue's multiplicity argument [SX91], or quasirandom groups [Gow08]. As $\mathsf{SL}_2(\mathbb{R})$ is not compact, such an argument can not be applied here. Instead, use the subgroup structure of $\mathsf{SL}_2(\mathbb{R})$, or in other words the two-transitivity of the Möbius action. Specifically, we show a "mixing property" for two-transitive actions.

Even using two-transitivity, proving a "mixing property" for the non-compact $\mathsf{SL}_2(\mathbb{R})$ is more difficult that in the finite/compact case. To complete the proof for $\mathsf{SL}_2(\mathbb{R})$, we need to use knowledge of the Fourier spectrum of the set A. We are able to obtain knowledge on the spectrum of A by adding to Ψ the translate map. The translate map implies that, w.l.o.g., we can assume that the spectrum of A does not have low frequencies.

3 A Monotone Expander

In essence, the maps defining the monotone expander are induced by the action of $\mathsf{SL}_2(\mathbb{R})$ on the projective real line. Consider the Möbius action: given $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ in $\mathsf{SL}_2(\mathbb{R})$, denote by $\overline{g}$ the map defined by

$$\overline{g}(x) = \frac{ax+b}{cx+d}.$$

These maps are monotone: for all g in $\mathsf{SL}_2(\mathbb{R})$, the derivative of the map $\overline{g}$ is

$$\overline{g}'(x) = \frac{1}{(cx+d)^2}.$$

So $\overline{g}$ is monotone in any interval not containing $-d/c$.

We are mostly interested in the restriction of $\overline{g}$ to the interval [0,1]. We also require that the maps we consider are defined over an interval. For these reasons, denote by $\tilde{g}$ the following map: If $-d/c$ is in $[0,1]$, define $\tilde{g}$ to be (say) the identity map. Otherwise (and more interestingly), $\tilde{g}$ is the restriction of $\overline{g}$ to inputs in the interval $[0,1] \cap \overline{g}^{-1}([0,1])$.

We describe two constructions of continuous monotone expanders, which we denote Ψ_0, Ψ. By construction, we shall have $\Psi \subset \Psi_0$. The reasons for the double definition are (i) Ψ_0 is easier to describe than Ψ but (ii) it is more natural to prove that Ψ is expanding (and hence so is Ψ_0).

For the constructions, we shall choose four numbers: small $\varepsilon > 0$ and large integers q, K, r.

The first construction. *Let Ψ_0 be the family of monotone smooth maps from subintervals of* [0,1] to [0,1] *defined as follows. For an integer k and a set of matrices $S \subset \mathsf{SL}_2(\mathbb{R})$, denote by $\mathcal{W}_k[S]$ the set of matrices that can be written as words of length at most k in elements of $S \cup S^{-1}$, where $S^{-1} = \{g^{-1} : g \in S\}$.* Let $\mathcal{G}_0 \subset \mathsf{SL}_2(\mathbb{R})$ be defined as

$$\mathcal{G}_0 = \mathcal{W}_1\left[\begin{pmatrix} 1 & 1/K \\ 0 & 1 \end{pmatrix}\right] \bigcup \mathcal{W}_r\left[\begin{pmatrix} 1 & 1/q \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1/q & 1 \end{pmatrix}\right].$$

Define

$$\Psi_0 = \{\tilde{g} : g \in \mathcal{G}_0\}.$$

The second construction (as we shall see) is a subset of the first construction. This (more elaborate) construction uses the following lemma. The proof of the lemma is given in Section 4.

Lemma 3. *There is a constant $C > 0$ so that the following holds. For $\varepsilon > 0$ small, there is a positive integer Q and a subset $\mathcal{G}$ of $\mathsf{SL}_2(\mathbb{R})$ so that*

1. $(1/\varepsilon)^{1/C} < Q < (1/\varepsilon)^C$,

2. $Q < |\mathcal{G}|^C$,
3. *elements of* $\mathcal{G}$ *freely generate a group,*
4. *elements of* $\mathcal{G}$ *have entries of the form* $\mathbb{Z}/Q$, *and*
5. *every* $g \in \mathcal{G}$ *admits*

$$\|g - 1\|_2 = (g_{1,1} - 1)^2 + (g_{1,2})^2 + (g_{2,1})^2 + (g_{2,2} - 1)^2 \leq \varepsilon.$$

The lemma summarizes all the properties $\mathcal{G}$ should satisfy in order to yield a monotone expander. When applying the lemma, ε is a small universal constant. An important (and useful) property of the lemma is that both $|\mathcal{G}|$ and Q are polynomially comparable to $1/\varepsilon$. Without this property, the lemma immediately follows from the strong Tits alternative [Bre08]. Property 4 yields the *non-commutative diophantine* property of $\mathcal{G}$, roughly, that for every $w \neq w'$ that are words of length k in elements of $\mathcal{G}$, the distance between w and w' is at least $(1/Q)^k$. This property is defined and used in [BG07]. Property 5 is crucial for handling the non-compactness of $\mathsf{SL}_2(\mathbb{R})$, and implies that if ε is small then for every g in $\mathcal{W}_1[\mathcal{G}]$, it holds that $-g_{2,2}/g_{2,1}$ is not in $[0,1]$, so $\tilde{g}$ is non-trivial.

The second construction. *Let* Ψ *be the family of monotone smooth maps* ψ *from sub-intervals of* $[0,1]$ *to* $[0,1]$ *defined as follows. Let* $\mathcal{G}$ *be the family of matrices given by Lemma 3 with respect to* ε. *Define*

$$\Psi = \left\{ \tilde{g} : g \in \mathcal{W}_1 \left[\mathcal{G} \cup \left\{ \begin{pmatrix} 1 & 1/K \\ 0 & 1 \end{pmatrix} \right\} \right] \right\}.$$

The following theorem shows that the two constructions indeed yield continuous monotone expanders.

Theorem 4. *There is a constant* $c_0 > 0$ *so that the following holds. Let* A *be a measurable subset of* $[0,1]$ *with* $|A| \leq 1/2$. *Then,*

$$|\Psi(A)| \geq (1 + c_0)|A|.$$

Specifically, $|\Psi_0(A)| \geq |\Psi(A)| \geq (1 + c_0)|A|$.

Theorem 4 implies Theorem 1, and follows from the following "restricted spectral gap" theorem. The Möbius action induces a unitary representation of $\mathsf{SL}_2(\mathbb{R})$ on $L^2(\mathbb{R})$ defined by

$$T_{g^{-1}} f(x) = \sqrt{\overline{g}'(x)} f(\overline{g}(x))$$

(the $\sqrt{\overline{g}'(x)}$ factors implies unitarity). For a positive integer K, denote by $\mathcal{F}_K$ the family of maps $f \in L^2(\mathbb{R})$ with $\mathsf{supp}(f) \subset [0,1]$ and $\|f\|_2 = 1$ so that for all $k \in \{1, 2, \ldots, K\}$,

$$\int_{I(k)} f(x)\, dx = 0,$$

where

$$I(k) = [(k-1)/K, k/K].$$

Theorem 5. *Let $\varepsilon > 0$ be a small enough constant. Let $\mathcal{G}$ be the set given by Lemma 3. If $K = K(\varepsilon)$ is a large enough positive integer, then for all $f \in \mathcal{F}_K$,*

$$\left\langle \sum_g \nu(g) T_g f, f \right\rangle < 1/2, \tag{1}$$

with the probability measure

$$\nu = (2|\mathcal{G}|)^{-1} \sum_{g \in \mathcal{G}} \mathbf{1}_g + \mathbf{1}_{g^{-1}},$$

where $\mathbf{1}_g$ is the delta function at g.

The "restricted spectral gap" theorem is proved in Section 5.

Proof of Theorem 4. We first reduce the general case to the "restricted spectral gap" case. Let $\sigma > 0$ be a small universal constant, to be determined. If there is an integer k between 1 and $K-1$ so that

$$\big||A \cap I(k+1)| - |A \cap I(k)|\big| \geq \sigma |A|,$$

then, using the three maps: $x \mapsto x + 1/K, x \mapsto x - 1/K$ and the identity map, that are defined by the three elements of $\mathcal{W}_1\left[\begin{pmatrix} 1 & 1/K \\ 0 & 1 \end{pmatrix}\right]$,

$$|\Psi(A)| \geq (1+\sigma)|A|.$$

It thus remains to consider the case that $\big||A \cap I(k+1)| - |A \cap I(k)|\big| < \sigma|A|$ for all k. Thus, for all k,

$$\big|K|A \cap I(k)| - |A|\big| < \sigma K^2 |A|. \tag{2}$$

Assume towards a contradiction that the theorem does not hold.

Since $\|g - 1\|_2 \leq \varepsilon$, for all $x \in [0,1]$,

$$\frac{1}{(1+2\varepsilon)^2} < \overline{g}'(x) < \frac{1}{(1-2\varepsilon)^2}. \tag{3}$$

Thus, for every $x \in [0,1]$,

$$0 \leq \overline{g}(x) - x < 10\varepsilon.$$

We replace $\tilde{g}$ by $\overline{g}$ by ensuring that even after applying maps in $\{\overline{g} : g \in \mathcal{W}_1[\mathcal{G}]\}$ we remain in $[0,1]$. To this end, let

$$A' = A \cap [k'/K, 1 - k'/K]$$

with k' the smallest integer so that $k' \geq 10\varepsilon K$. By (2),

$$0.99|A| \leq |A'| \leq |A|, \tag{4}$$

as long as σ, ε are small. Denote

$$f = \mathbf{1}_{A'} - |A'|.$$

Project A' on $\mathcal{F}_K$. Define F as follows: for all $x \in [0,1]$, if $x \in I(k)$, then

$$F(x) = \mathbf{1}_{A'}(x) - K|A' \cap I(k)|. \quad \text{Hence,} \quad F/\left\|F\right\|_2 \in \mathcal{F}_K.$$

Our goal is to use F and assumption to obtain a contradiction to Theorem 5.

Claim 6. *First,*

$$\|f - F\|_2^2 \leq 0.01 \left\|F\right\|_2^2.$$

Second, for every $g \in \mathcal{G} \cup \mathcal{G}^{-1}$,

$$\langle T_g f, f \rangle \geq 0.8 \left\|F\right\|_2^2.$$

Before proving the claim, we show how it completes the proof. By averaging,

$$\begin{aligned} 0.8 \left\|F\right\|_2^2 \leq \sum_g \nu(g) \langle T_g f, f \rangle &= \left\langle \sum_g \nu(g) T_g(f - F + F), f - F + F \right\rangle \\ &\leq 3 \|f - F\|_2^2 + \left\langle \sum_g \nu(g) T_g F, F \right\rangle \leq 0.1 \left\|F\right\|_2^2 + \left\langle \sum_g \nu(g) T_g F, F \right\rangle. \end{aligned}$$

This contradicts Theorem 5.

Proof of Claim 6. First, for σ small, using (2) and (4),

$$\begin{aligned} \|f - F\|_2^2 &= \sum_{k=k'}^{K-k'} \int_{I(k)} (K|A' \cap I(k)| - |A'|)^2 \, dx \leq \sum_{k=k'}^{K-k'} \int_{I(k)} (|A| - |A'| + \sigma K^2 |A|)^2 \, dx \\ &\leq \sum_{k=k'}^{K-k'} \int_{I(k)} (0.01 + \sigma K^2)^2 |A|^2 \, dx \leq 0.001 |A|^2. \end{aligned}$$

So, using (4) again, since $|A| \leq 1/2$,

$$0.99|A| \leq \sqrt{|A'|(1 - |A'|)} = \left\|f\right\|_2 \leq \|f - F\|_2 + \left\|F\right\|_2 \leq 0.1|A| + \left\|F\right\|_2.$$

This implies the first part of claim.

Second, Equation (3) and unitarity of $T_{g^{-1}}$ imply

$$(1 - 5\varepsilon)|A'| = (1 - 5\varepsilon) \left\|\mathbf{1}_{A'}\right\|_2^2 = (1 - 5\varepsilon) \left\|T_{g^{-1}} \mathbf{1}_{A'}\right\|_2^2 \leq |\overline{g}^{-1}(A')|.$$

Assumption that expansion fails thus implies

$$|\overline{g}^{-1}(A') \cap A'| = |\overline{g}^{-1}(A')| - |\overline{g}^{-1}(A') \setminus A'| \geq (1-5\varepsilon)|A'| - c_0|A'|.$$

Since $|A'| \leq 1/2$ and by (3), therefore,

$$\begin{aligned}\left\langle T_{g^{-1}} f, f\right\rangle &= \left\langle T_{g^{-1}} \mathbf{1}_{A'}, \mathbf{1}_{A'}\right\rangle - \left\langle \mathbf{1}_{A'}, T_g |A'|\right\rangle - \left\langle T_{g^{-1}}|A'|, \mathbf{1}_{A'}\right\rangle + \left\langle T_{g^{-1}}|A'|, |A'|\right\rangle \\ &\geq (1-5\varepsilon)|\overline{g}^{-1}(A') \cap A'| - (1+5\varepsilon)|A'||A'| - (1+5\varepsilon)|A'||A'| \\ &\quad +(1-5\varepsilon)|A'||A'| \\ &\geq (1-5\varepsilon)(1-5\varepsilon-c_0)|A'| - (1+15\varepsilon)|A'||A'| \\ &\geq 0.9|A'|(1-|A'|),\end{aligned}$$

as long as ε, c_0 are small. Finally, using (2), for σ small,

$$\begin{aligned}\|F\|_2^2 &= \sum_{k=k'}^{K-k'} |A' \cap I(k)|(1 - K|A' \cap I(k)|) \\ &< \sum_{k=k'}^{K-k'} |A' \cap I(k)|(1 - |A'|(1-\sigma K^2)) \leq 1.01|A'|(1-|A'|).\end{aligned}$$

□

4 The Second Construction: Finding Set of Generators

For $\delta > 0$, denote by $B_\delta(x)$ the ball of radius δ around x and by $\Gamma_\delta(A)$ the δ-neighborhood of the set A. We consider the L^2-metric on $\mathsf{SL}_2(\mathbb{R})$.

Proof of Lemma 3. Breuillard [Bre08] proved a strong Tits alternative: there is a constant $r \in \mathbb{Z}$ so that if S is a finite symmetric subset of $\mathsf{SL}_2(\mathbb{R})$, which generates a non-amenable subgroup, then $S_{(r)} = \{s_1 s_2 \dots s_r : s_i \in S\}$ contains two elements that freely generate a group.

Let

$$h_1 = \begin{pmatrix} 1 & 1/q \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad h_2 = \begin{pmatrix} 1 & 0 \\ 1/q & 1 \end{pmatrix}.$$

Observe

$$h_1^{2q} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad h_2^{2q} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix},$$

which are known to generate a free group. Hence, h_1, h_2 generate a non-amenable group. Apply the strong Tits alternative on the set $S = \{h_1, h_2, h_1^{-1}, h_2^{-1}\}$. There are thus $g_1, g_2 \in S_{(r)}$ that freely generate a group.

It remains to convert g_1, g_2 to many elements that are close to identity and freely generate a group. Let $\ell \sim \log(1/\varepsilon)$ so that the following holds. Consider

$$W = \left\{w^2 : \ w = s_1 \dots s_\ell, \ s_1 = g_1, \ s_\ell = g_2, \ s_i \in \{g_1, g_2, g_1^{-1}, g_2^{-1}\}, \ s_{i+1} \neq s_i^{-1}\right\}.$$

Say that a word $\sigma_1\sigma_2\dots\sigma_k$ in an alphabet $\Sigma\cup\Sigma^{-1}$ is $\langle\Sigma\rangle$-*reduced* if $\sigma_{i+1}\neq\sigma_i^{-1}$ for all $i\in\{1,\dots,k-1\}$. The size of W is order 3^ℓ and W consists of words of $\langle g_1,g_2\rangle$-reduced-length exactly 2ℓ.

Claim 7. *The elements of W freely generate a group.*

Proof. Let $w_1\neq w_2^{-1}$ in $W\cup W^{-1}$. Write

$$w_1=(g_{a_1}s_1g_{b_1})^2 \quad\text{and}\quad w_2=(g_{a_2}s_2g_{b_2})^2$$

with s_1,s_2 reduced words in $\langle g_1,g_2\rangle$, and $g_{a_1},g_{b_1},g_{a_2},g_{b_2}$ in $\{g_1,g_2,g_1^{-1},g_2^{-1}\}$. If either $w_1,w_2\in W$ or $w_1,w_2\in W^{-1}$, then $g_{a_2}\neq g_{b_1}^{-1}$ and so

$$w_1w_2=g_{a_1}s_1g_{b_1}g_{a_1}s_1g_{b_1}g_{a_2}s_2g_{b_2}g_{a_2}s_2g_{b_2}$$

in $\langle g_1,g_2\rangle$-reduced form. If either $w_1\in W,w_2\in W^{-1}$ or $w_1\in W^{-1},w_2\in W$, then, since $s_1\neq s_2^{-1}$ and the reduced-length of both s_1,s_2 is $\ell-2$,

$$w_1w_2=g_{a_1}s_1g_{b_1}g_{a_1}s'g_{b_2}g_{a_2}s_2g_{b_2}$$

in $\langle g_1,g_2\rangle$-reduced form, with s' non-trivial.

Any non-trivial $\langle W\rangle$-reduced word is not the identity of $\langle g_1,g_2\rangle$: for $w=g_aszsg_b$ in $W\cup W^{-1}$, where z is a product of two elements of $\{g_1,g_2,g_1^{-1},g_2^{-1}\}$, call z the *center* of w. The above implies that if $w_1\neq w_2^{-1}$ then the centers of both w_1,w_2 are not reduced in the $\langle g_1,g_2\rangle$-reduced form of w_1w_2.

Hence, if $w=w_1w_2\dots w_k$ is a non-trivial $\langle W\rangle$-reduced word, then even in its $\langle g_1,g_2\rangle$-reduced form w is not the identity (as all centers are not reduced). □

Observe that for every $w\in W$,

$$\|w\|_2,\left\|w^{-1}\right\|_2\leq(1+1/q)^{2r\ell}:=N.$$

Cover the ball $B_N(1)$ in $\mathsf{SL}_2(\mathbb{R})$ with balls of radius ε/N. There exists $w_0\in W$ so that

$$\left|B_{\varepsilon/N}(w_0)\cap W\right|\gtrsim|W|(\varepsilon/N^2)^3\gtrsim\varepsilon^3 3^\ell(1+1/q)^{-12r\ell}.$$

Define

$$\mathcal{G}=\left(w_0^{-1}\left(B_{\varepsilon/N}(w_0)\cap W\right)\right)\setminus\{1\}.$$

Choose q as a universal constant so that $(1+1/q)^{12r}<1.01$. Hence,

$$|\mathcal{G}|\gtrsim\varepsilon^3 3^\ell(1+1/q)^{-12r\ell}-1\gtrsim 2^\ell.$$

In addition, for $g\in\mathcal{G}$,

$$\|1-g\|_2\leq N\|w_0-w_0g\|_2\leq\varepsilon,$$

and the entries of g are of the form $\mathbb{Z}/Q$ with $Q=q^{4r\ell}$ and $\log Q\sim\log(1/\varepsilon)$. Finally, as $\mathcal{G}$ is of the form $w_0^{-1}W\setminus\{1\}$ with W freely generating a group, the elements of $\mathcal{G}$ freely generate a group as well.

5 Restricted Spectral Gap Via Flattening

To prove the "restricted spectral gap" property, we prove the following theorem that roughly states that after enough iterations ν becomes very flat. Denote by P_δ the *approximate identity* on $\mathsf{SL}_2(\mathbb{R})$, namely, the density of the uniform distribution on the ball of radius δ around 1 in $\mathsf{SL}_2(\mathbb{R})$,

$$P_\delta = \frac{\mathbf{1}_{B_\delta(1)}}{|B_\delta(1)|}.$$

For two distributions μ, μ' on $\mathsf{SL}_2(\mathbb{R})$ denote by $\mu * \mu'$ the convolution of μ and μ'. Denote by $\mu^{(\ell)}$ the ℓ-fold convolution of μ with itself.

Theorem 8. *Let $\gamma > 0$. Assume that $\varepsilon > 0$, the parameter from 5 in Lemma 3, and $\delta > 0$ are small enough as a function of γ. If*

$$\ell > C_1 \frac{\log(1/\delta)}{\log(1/\varepsilon)}$$

with $C_1 = C_1(\gamma) > 0$, then

$$\left\| \nu^{(\ell)} * P_\delta \right\|_\infty < \delta^{-\gamma}.$$

The proof of the theorem is given in Section 6 (when applying the theorem, γ is a universal constant).

Proof of Theorem 5. Let $f \in \mathcal{F}_K$. Assume that (1) does not hold, i.e.,

$$\langle Tf, f \rangle \geq 1/2, \tag{5}$$

where

$$T = \sum_g \nu(g) T_g.$$

We start by finding a level set of the Fourier transform that "violates (1) as well." The Littlewood–Paley decomposition of f is

$$f = \sum_k \Delta_k f,$$

where for every integer k and $2^k \leq |\lambda| < 2^{k+1}$ for every $\lambda \in \operatorname{supp} \widehat{\Delta_k f}$.

As $f \in \mathcal{F}_K$, we can consider the part of f with high frequencies.

Claim 9. *For $k_0 \geq 0$, define*

$$f_0 = \sum_{k \geq k_0} \Delta_k f.$$

If K is large enough, depending on k_0, then

$$\langle Tf_0, f_0 \rangle > 1/4.$$

Proof. For every $\lambda \in \mathbb{R}$,

$$\widehat{f}(\lambda) = \sum_{k'=1}^{K} \int_0^{1/K} f(y_{k'} + x) e^{-2\pi i \lambda (y_{k'} + x)} \, dx.$$

where $y_{k'} = 1 + (k' - 1)/K$. For fixed k', since the integral of f over the interval $I(k) = [y_{k'}, y_{k'} + 1/K]$ is zero,

$$\left| \int_0^{1/K} f(y_{k'} + x) e^{2\pi i \lambda (y_{k'} + x)} \, dx \right| \lesssim \frac{|\lambda|}{K} \int_0^{1/K} |f(y_{k'} + x)| dx.$$

Hence, for every integer k,

$$\|\Delta_k f\|_2 \lesssim 2^k / K.$$

So,

$$\|f - f_0\|_2 \lesssim 2^{k_0}/K \leq 1/20$$

for K large. Thus,

$$1/2 < \langle Tf, f \rangle \leq \langle Tf_0, f_0 \rangle + 3 \|f - f_0\|_2 < \langle Tf_0, f_0 \rangle + 1/4.$$

Isolate one frequency-level of f_0, using the following claim. □

Claim 10. *There is $k \geq k_0$ so that*

$$\|T\Delta_k f_0\|_2 \geq c_1 \|\Delta_k f_0\|_2$$

with $c_1 > 0$ a universal constant.

Proof. Bound

$$\begin{aligned} \|Tf_0\|_2^2 &\leq \sum_{k,k'} |\langle T\Delta_k f_0, T\Delta_{k'} f_0 \rangle| \\ &= \sum_{|k-k'| \leq C} |\langle T\Delta_k f_0, T\Delta_{k'} f_0 \rangle| + \sum_{|k-k'| > C} |\langle T\Delta_k f_0, T\Delta_{k'} f_0 \rangle| \end{aligned}$$

with $C > 0$ a universal constant to be determined. Bound each of the two terms in the sum separately. First,

$$\sum_{|k-k'| \leq C} |\langle T\Delta_k f_0, T\Delta_{k'} f_0 \rangle| \leq \sum_{|k-k'| \leq C} \|T\Delta_k f_0\|_2 \|T\Delta_{k'} f_0\|_2 \lesssim C \sum_k \|T\Delta_k f_0\|_2^2 .$$

Second, consider $k > k' + C$. The (absolute value of the) spectrum of $\Delta_k f_0$ is order 2^k, and of $\Delta_{k'} f_0$ is order $2^{k'}$. The operator T_g for $g \in (\mathcal{G} \cup \mathcal{G}^{-1})(\mathcal{G} \cup \mathcal{G}^{-1})$ is a smooth L^∞-perturbation of identity. Hence, for some $g \in (\mathcal{G} \cup \mathcal{G}^{-1})(\mathcal{G} \cup \mathcal{G}^{-1})$,

$$\begin{aligned}
|\langle T\Delta_k f_0, T\Delta_{k'} f_0\rangle| &\leq |\langle \Delta_k f_0, T_g \Delta_{k'} f_0\rangle| \\
&\sim \left| \int\limits_{\lambda \sim 2^k} \frac{\widehat{\Delta_k f_0}(\lambda)}{\lambda} \cdot \lambda \, \widehat{(T_g \Delta_{k'} f_0)}(\lambda)\, d\lambda \right| \\
&\lesssim 2^{-k} \left\|\Delta_k f_0\right\|_2 \cdot \left\|(T_g \Delta_{k'} f_0)'\right\|_2 \\
&\lesssim 2^{-k} \left\|\Delta_k f_0\right\|_2 2^{k'} \left\|\Delta_{k'} f_0\right\|_2 .
\end{aligned}$$

Thus,

$$\sum_{k > k' + C} |\langle T\Delta_k f_0, T\Delta_{k'} f_0\rangle| \lesssim \sum_{k > k' + C} 2^{k' - k} \left\|\Delta_k f_0\right\|_2 \left\|\Delta_{k'} f_0\right\|_2 \lesssim 2^{-C} \left\|f_0\right\|_2^2 ,$$

and so, for appropriate C,

$$\sum_{|k - k'| > C} |\langle T\Delta_k f_0, T\Delta_{k'} f_0\rangle| < 1/20.$$

Concluding, using Claim 9,

$$\sum_{k \geq k_0} \left\|\Delta_k f_0\right\|_2^2 \lesssim \left\|f_0\right\|_2^2 \lesssim 1/16 - 1/20 < \left\|Tf_0\right\|_2^2 - 1/20 \lesssim C \sum_{k \geq k_0} \left\|T\Delta_k f_0\right\|_2^2 . \qquad \square$$

Set

$$F = \frac{\Delta_k f_0}{\left\|\Delta_k f_0\right\|_2}$$

with k from Claim 10. Thus, $\langle TF, TF\rangle \geq c_1^2$ and so $\left\|T^2 F\right\|_2 \geq c_1^2$. Iterating, for all $\ell > 0$ a power of two,

$$c_1^\ell \leq \left\|T^{\ell/2} F\right\|_2^2 = \left\langle F, T^\ell F\right\rangle \leq \left\|T^\ell F\right\|_2 . \tag{6}$$

To prove the theorem, argue that the norm of $T^\ell F$ is actually small, thus obtaining the required contradiction: let $\gamma > 0$ be a small universal constant (to be determined). Let ℓ be the smallest power of two so that

$$\ell > C_1(\gamma) k / \log(1/\varepsilon)$$

and by Theorem 8,

$$\left\|\nu^{(\ell)} * P_\delta\right\|_\infty < \delta^{-\gamma},$$

with $\varepsilon > 0$ a small enough universal constant to be determined, and

$$\delta = 4^{-k}.$$

As δ is small and the spectrum of F is controlled, the following claim holds.

Claim 11.

$$\left\| \int\limits_{\mathsf{SL}_2(\mathbb{R})} (T_g F)((\nu^{(\ell)} * P_\delta)(g))\, dg \right\|_2 \gtrsim c_1^\ell.$$

Proof. If $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ satisfies $\|g-1\|_2 \leq \eta \leq 1/20$, then for all $x \in \mathbb{R}$ so that $|x| \leq 2$,

$$|x - gx| = \left| \frac{cx^2 + dx - ax - b}{cx + d} \right| \lesssim \eta.$$

In addition, if $h \in B_\delta(g)$ for $g \in \mathsf{supp}(\nu^{(\ell)})$, then

$$\left\|h^{-1}g - 1\right\|_2 \leq \delta(1+\varepsilon)^\ell.$$

Recall, $2^k\delta(1+\varepsilon)^\ell$ is much smaller than c_1^ℓ. Hence, since the norm of the derivative of F is at most order 2^k,

$$\|T_g F - T_h F\|_2 = \left\|F - T_{h^{-1}g} F\right\|_2 \lesssim 2^k \delta (1+\varepsilon)^\ell.$$

So,

$$\left\| T^\ell F - \int\limits_{\mathsf{SL}_2(\mathbb{R})} (T_h F)((\nu^{(\ell)} * P_\delta)(h))\, dh \right\|_2 \lesssim 2^k (1+\varepsilon)^\ell \delta \leq c_1^\ell/2.$$

The claim follows by (6). □

The claim above contradicts the following proposition, as shown below. In short, the proposition follows by the flatness lemma and the subgroup structure of $\mathsf{SL}_2(\mathbb{R})$.

Proposition 12. *There exists universal constants $\sigma_0, C > 0$ so that*

$$\left\| \int\limits_{\mathsf{SL}_2(\mathbb{R})} (T_g F)((\nu^{(\ell)} * P_\delta)(g))\, dg \right\|_2 \lesssim \delta^{-\gamma} (1+\varepsilon)^{C\ell} 2^{-\sigma_0 k}.$$

Proof. Bound, using Theorem 8 and unitarity of T_h, since the support of $\nu^{(\ell)} * P_\delta$ is contained in $B_{2(1+\varepsilon)^\ell}(1)$,

$$\begin{aligned} \left\| \int (T_g F)((\nu^{(\ell)} * P_\delta)(g))\, dg \right\|_2^2 &= \int\int \langle T_g F, T_h F \rangle\, ((\nu^{(\ell)} * P_\delta)(g))((\nu^{(\ell)} * P_\delta)(h))\, dg\, dh \\ &\lesssim \delta^{-2\gamma}(1+\varepsilon)^{3\ell} \int\limits_{B_{4(1+\varepsilon)^{2\ell}}(1)} |\langle T_g F, F \rangle|\, dg. \end{aligned} \tag{7}$$

Approximate $B_{4(1+\varepsilon)^{2\ell}}(1)$ by a smooth function: let $\kappa : \mathsf{SL}_2(\mathbb{R}) \to \mathbb{R}_{\geq 0}$ be a smooth function so that $\|\kappa\|_\infty = 1$, and so that $\kappa(g) = 1$ if $\|g-1\|_2 \leq 4(1+\varepsilon)^{2\ell}$ and $\kappa(g) = 0$ if $\|g\|_2 > 8(1+\varepsilon)^{2\ell}$. Using Cauchy–Schwarz inequality,

$$|(7)| \lesssim \delta^{-2\gamma}(1+\varepsilon)^{5\ell} \left(\int |\langle T_g F, F\rangle|^2 \, \kappa(g)\, dg \right)^{1/2}. \tag{8}$$

Write

$$\int |\langle T_g F, F\rangle|^2 \, \kappa(g)\, dg \leq \int\int |F(x)||F(y)| \left| \int T_g F(x) T_g F(y) \kappa(g)\, dg \right| dx\, dy.$$

Separate to two cases, according to the distance between x and y. Choose $\eta > 0$ small, to be determined. In both cases, use the following (convenient) parameterization of $\mathsf{SL}_2(\mathbb{R})$:

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} u\cos\theta & v\cos\phi \\ u\sin\theta & v\sin\phi \end{pmatrix}$$

with

$$uv\sin(\phi-\theta) = 1.$$

On the chart $a \neq 0$, we have

$$dg = \frac{da\, db\, dc}{|a|} = \frac{du\, d\theta\, d\phi}{|u|\sin^2(\theta-\phi)}.$$

Case one. The first case is when x, y are close: bound

$$\int\limits_{|x-y|<\eta} \int |F(x)||F(y)| \int |T_g F(x)||T_g F(y)| \kappa(g)\, dg\, dx\, dy. \tag{9}$$

Write $F = F_1 + F_\infty$ with

$$\|F_1\|_1 \leq 2^{-\sigma k} \quad \text{and} \quad \|F_\infty\|_\infty \leq 2^{\sigma k}$$

for a universal constant $\sigma > 0$ to be determined. Equation (9) can be bounded from above by a sum of several terms (with different combinations of F_1, F_∞ replacing F). Consider, e.g., substituting F_1 instead of the leftmost F in (9),

$$\begin{aligned} &\int\limits_{|x-y|<\eta} \int |F_1(x)||F(y)| \int |T_g F(x)||T_g F(y)| \kappa(g)\, dg\, dx\, dy \\ &\leq \int |F_1(x)| \int |T_g F(x)| \kappa(g)\, dg\, dx. \end{aligned} \tag{10}$$

Fix x, and denote

$$M = (x+1)^{-1/2} \begin{pmatrix} 1 & -x \\ 1 & 1 \end{pmatrix} \in \mathsf{SL}_2(\mathbb{R}),$$

so that $\overline{M}(x) = 0$ (the matrix M shows two-transitivity of the Möbius action: M maps x to zero and -1 to infinity. Note that $x, -1$ are far). Change variables and use parametrization given above,

$$\int |T_g F(x)|\kappa(g)\,dg = \int |T_{M^{-1}g^{-1}} F(x)|\kappa(M^{-1}g^{-1})\,dg$$
$$\lesssim \iiint |F(\cot\phi)|\kappa(M^{-1}g^{-1})\frac{1}{|\sin\phi||\sin(\theta-\phi)|}\,du\,d\theta\,d\phi. \quad (11)$$

If $\kappa(M^{-1}g^{-1}) \neq 0$, then $\|g\|_2 \lesssim (1+\varepsilon)^{2\ell}$, and so in the integral above $|\sin(\theta-\phi)| \gtrsim (1+\varepsilon)^{-4\ell}$.

$$|(11)| \lesssim (1+\varepsilon)^{4\ell} \int\int\int |F(\xi)|\kappa(M^{-1}g^{-1})\frac{1}{|\xi+1|^{1/2}}\,du\,d\theta\,d\xi \lesssim (1+\varepsilon)^{6\ell}.$$

Hence,

$$|(10)| \lesssim (1+\varepsilon)^{6\ell}\,\|F_1\|_1 \leq (1+\varepsilon)^{6\ell}2^{-\sigma k}.$$

The same bound holds also if we replace each of the other three F's by F_1 in (9). It thus remains to trivially bound

$$\int \int_{|x-y|<\eta} |F_\infty(x)||F_\infty(y)| \int |T_g F_\infty(x)||T_g F_\infty(y)|\kappa(g)\,dg\,dx\,dy \lesssim \eta(1+\varepsilon)^{6\ell}2^{4\sigma k},$$

and conclude

$$|(9)| \lesssim (1+\varepsilon)^{6\ell}\left(\eta 2^{4\sigma k} + 2^{-\sigma k}\right). \quad (12)$$

Case two. Next, understand what happens for far x and y. The argument in this case is more elaborate and uses knowledge of the spectrum of F. Start by

$$\int \int_{|x-y|\geq\eta} |F(x)||F(y)| \left|\int T_g F(x) T_g F(y)\kappa(g)\,dg\right| dx\,dy$$
$$\leq \left(\int \int_{|x-y|\geq\eta} \left|\int_{\mathsf{SL}_2(\mathbb{R})} T_g F(x) T_g F(y)\kappa(g)\,dg\right|^2 dx\,dy\right)^{1/2}. \quad (13)$$

In this case, argue for fixed x and y in $[0,1]$ so that $x \geq y+\eta$. Denote

$$M = (x-y)^{-1/2}\begin{pmatrix} 1 & -x \\ 1 & -y \end{pmatrix} \in \mathsf{SL}_2(\mathbb{R}),$$

so that $\overline{M}(x) = 0$ and $\overline{M}(y) = \infty$. Change variables,

$$\left|\int T_g F(x) T_g F(y)\kappa(g)\,dg\right| = \left|\int T_{M^{-1}g^{-1}} F(x) T_{M^{-1}g^{-1}} F(y)\kappa(M^{-1}g^{-1})\,dg\right|$$
$$= (x-y)^{-1}\left|\int \frac{F(\cot\phi)F(\cot\theta)}{|\sin\phi\cdot\sin\theta|}\kappa(M^{-1}g^{-1})\frac{du\,d\theta\,d\phi}{|u||\sin(\theta-\phi)|}\right|.$$

Change variables,

$$\int \frac{F(\cot\phi)F(\cot\theta)}{|\sin\phi\cdot\sin\theta|}\kappa(M^{-1}g^{-1})\frac{du\,d\theta\,d\phi}{|u||\sin(\theta-\phi)|} = \int\int F(\xi)F(\zeta)E(\xi,\zeta)d\xi\,d\zeta,$$

with

$$E(\xi,\zeta) = \frac{\sqrt{(1+\xi^2)(1+\zeta^2)}}{|\sin(\cot^{-1}\zeta-\cot^{-1}\xi)|}\int \kappa(M^{-1}g^{-1})\frac{du}{|u|}.$$

Continue by using that Fourier basis diagonalize ∇. Start by bounding the norms of E and ∇E. First, if $\kappa(M^{-1}g^{-1})\neq 0$, then

$$\|g\|_2 \lesssim (1+\varepsilon)^{2\ell}\eta^{-1/2}.$$

Hence, in the definition of E we can assume

$$(1+\varepsilon)^{-2\ell}\eta^{1/2} \lesssim |u| \lesssim (1+\varepsilon)^{2\ell}\eta^{-1/2},$$

and

$$\frac{1}{|\sin(\cot^{-1}\zeta-\cot^{-1}\xi)|} \gtrsim (1+\varepsilon)^{-4\ell}\eta.$$

Therefore, there is a universal constant $C>0$ so that

$$\|E\|_\infty, \left\|\|\nabla E\|_2\right\|_\infty \lesssim (1+\varepsilon)^{C\ell}\eta^{-C}.$$

Since the support of the Fourier transform of F is of absolute value at least order 2^k, bound

$$\left|\int\int F(z)F(w)E(z,w)\,dz\,dw\right| \lesssim 2^{-k}(1+\varepsilon)^{C\ell}\eta^{-C}.$$

Thus,

$$|(13)| \leq 2^{-k}(1+\varepsilon)^{C\ell}\eta^{-C-1}. \tag{14}$$

Concluding. By (12) and (14),

$$\sqrt{|(7)|} \lesssim \delta^{-\gamma}(1+\varepsilon)^{C\ell}\left(\eta 2^{4\sigma k}+2^{-\sigma k}+2^{-k}\eta^{-C}\right)^{1/2} \leq \delta^{-\gamma}(1+\varepsilon)^{C\ell}2^{-\sigma k/4},$$

for appropriate choice of η, and with $\sigma>0$ a universal constant. □

We can finally conclude, using Claim 11 and Proposition 12,

$$c_1^\ell \lesssim |(8)| \lesssim \delta^{-\gamma}(1+\varepsilon)^{C\ell}2^{-\sigma_0 k}, \tag{15}$$

which is a contradiction for $\gamma=\sigma_0/4$, k_0 large and ε small. □

6 Flatness Via a Product Theorem

Theorem 8 follows from the following flattening lemma, which roughly states that if

$$\mu = \nu^{(\ell_0)} * P_\delta$$

is a little flat then $\mu * \mu$ is much flatter (unless μ is already very flat). The proof of the lemma is given in Section 6.1.

LEMMA 13. *Let $0 < \gamma < 3/2$. With the notation above, assume that*

$$\delta^{-\gamma} < \|\mu\|_2 < \delta^{-3/2+\gamma}$$

and

$$\ell_0 > C_2 \frac{\log(1/\delta)}{\log(1/\varepsilon)}$$

with $C_2 = C_2(\gamma) > 0$. Also assume that $\varepsilon > 0$, the parameter from 5 in Lemma 3, and $\delta > 0$ are small enough as a function of γ. Then, there exists $\sigma = \sigma(\gamma) > 0$ so that

$$\|\mu * \mu\|_2 < \delta^\sigma \|\mu\|_2 .$$

We apply the flattening lemma iteratively. To start iterating, we need to show that μ is "a little flat" to begin with.

PROPOSITION 14. *If*

$$\ell_0 \geq \log_Q(1/\delta)$$

with Q from Lemma 3, then

$$\|\mu\|_2 \leq \delta^{-3/2+\gamma}$$

with $\gamma > 0$ a universal constant.

This follows from Kesten's bound [Kes59], the following proposition about random walks on free groups.

PROPOSITION 15. *Assume H is a finite set freely generating a group $\langle H \rangle$. Denote*

$$\pi = (2|H|)^{-1} \sum_{h \in H} \mathbf{1}_h + \mathbf{1}_{h^{-1}},$$

a probability distribution on $\langle H \rangle$. Denote by $p^{(t)}(x,x)$ the probability of being at x after t steps of a random walk on the group $\langle H \rangle$ according to π started at x. Then,

$$\limsup_{t \to \infty} (p^{(t)}(x,x))^{1/t} = \frac{\sqrt{2|H|-1}}{|H|}.$$

Proof of Proposition 14. Let k be the maximal integer so that

$$1/Q^k \geq \delta^{1/2}.$$

For every $y \in \mathsf{supp}(\nu^{(k)})$,

$$\|y\|_2 \leq (1+\varepsilon)^k \leq \delta^{-\varepsilon},$$

for ε small. By Lemma 3, the entries of elements in $\mathcal{W}_k(\mathcal{G})$ are in $\mathbb{Z}/Q^k$. So, for all $y \neq y'$ in $\mathcal{W}_k(\mathcal{G})$,

$$\left\|y - y'\right\|_2 \geq \delta^{1/2},$$

which implies

$$(yB_\delta(1)) \cap (y'B_\delta(1)) = \emptyset,$$

for ε small. Hence,

$$\left\|\sum_y \nu^{(k)}(y) P_\delta(y^{-1}\cdot)\right\|_2 \leq \left(\sum_y (\nu^{(k)}(y))^2 \left\|P_\delta(y^{-1}\cdot)\right\|_2^2\right)^{1/2} \leq \left\|\nu^{(k)}\right\|_\infty^{1/2} \|P_\delta\|_2 .$$

Finally, by Proposition 15 and Lemma 3, since convolution does not increase norms,

$$\|\mu\|_2 \lesssim \left(\frac{2|\mathcal{G}|-1}{|\mathcal{G}|^2}\right)^{k/4} \delta^{-3/2} < \delta^{-3/2+\gamma}. \qquad \square$$

Proof of Theorem 8. By Proposition 14, and Lemmas 3 and 13,

$$\left\|\mu^{(k)}\right\|_2 = \left\|(\nu^{(\ell_0)} * P_\delta)^{(k)}\right\|_2 \leq \delta^{-\gamma/4} \tag{16}$$

with $k = k(\gamma) > 1$ and

$$\ell_0 \leq C_3 \frac{\log(1/\delta)}{\log(1/\varepsilon)},$$

with $C_3 > 0$ a constant. For every g,

$$\left|\mu^{(2k)}(g)\right| = \left|\int_h \mu^{(k)}(h)\mu^{(k)}(h^{-1}g)\, dh\right| \leq \left\|\mu^{(k)}\right\|_2^2 \leq \delta^{-\gamma/2}.$$

Lemma 2.5 in [BG07] states (for $\mathsf{SU}(2)$ but the same proof holds in our case)

$$cP_\delta \leq P_\delta * P_\delta \leq \frac{1}{c} P_{2\delta}$$

with $c > 0$ a constant. Hence,

$$\left\|\nu^{(\ell)} * P_\delta\right\|_\infty \leq C_4(1+\varepsilon)^{C_4\ell_0} \left\|\mu^{(2k)}\right\|_\infty \leq C_4(1+\varepsilon)^{C_4\ell_0}\delta^{-\gamma/2} \leq \delta^{-\gamma}$$

with $C_4 = C_4(\gamma) > 0$ and $\ell \leq C_4\ell_0$, for ε, δ small. $\square$

6.1 A product theorem. The flattening lemma follows from the following product theorem (the proof of the product theorem is deferred to Section 7). We need to use *metric entropy*: for a subset S of a metric space denote by $\mathcal{N}_\delta(S)$ the least number of balls of radius δ needed to cover S.

Theorem 16. *For all $\sigma_1, \tau > 0$, there is $\varepsilon_5 > 0$ so that the following holds. Let $\delta > 0$ be small enough. Let $A \subset \mathsf{SL}_2(\mathbb{R}) \cap B_\alpha(1), \alpha > 0$ a small universal constant, be so that*

1. *$A = A^{-1}$,*
2. *$\mathcal{N}_\delta(A) = \delta^{-3+\sigma_0}$, $\sigma_1 \leq \sigma_0 \leq 3 - \sigma_1$,*
3. *for every $\delta < \rho < \delta^{\varepsilon_5}$, there is a finite set $X \subset A$ so that $|X| \geq \rho^{-\tau}$ and for every $x \neq x'$ in X we have $\|x - x'\|_2 \geq \rho$, and*
4. *w.r.t. every complex basis change diagonalizing some matrix in $\mathsf{SL}_2(\mathbb{R}) \cap B_1(1)$, there is $g \in A_{(4)}$ so that $|g_{1,2} g_{2,1}| \geq \delta^{\varepsilon_5}$.*

Then,

$$\mathcal{N}_\delta(AAA) > \delta^{-\varepsilon_5} \mathcal{N}_\delta(A).$$

The condition that A is contained in a small ball is not necessary, but simplifies the statement and proof. The condition $A = A^{-1}$ is, of course, not necessary as well, but simplifies notation. Condition 4 above implies that A is far from strict subgroups.

Proof of Lemma 13. We prove the lemma for

$$\ell_0 \sim C_2(\gamma) \frac{\log(1/\delta)}{\log(1/\varepsilon)}.$$

The proof for larger ℓ_0 follows, as convolution does not increase the norm.

Assume towards a contradiction that

$$\|\mu * \mu\|_2 > \delta^\sigma \|\mu\|_2 .$$

To prove the theorem, we shall find a set A that violates the product theorem. The set A will be one of the level sets of μ in the following decomposition. Decompose μ as

$$\mu \sim \sum_j 2^j \chi_j,$$

where the sum is over $O(\log(1/\delta))$ values of j (recall that μ is point-wise bounded by $O(1/\delta^3)$ and we can ignore points with too small μ-measure), and where χ_j is the characteristic function of a set $A_j \subset \mathsf{SL}_2(\mathbb{R})$ so that

$$A_j = A_j^{-1}. \tag{17}$$

Choose $j_1 < j_2$ so that

$$2^{j_1+j_2} \|\chi_{j_1} * \chi_{j_2}\|_2 \gtrsim \|\mu * \mu\|_2 / \log^2(1/\delta) \geq \delta^{0+} \|\mu\|_2 . \tag{18}$$

Using Young's inequality, bound

$$2^{j_1+j_2}\,\|\chi_{j_1}\|_2\,\|\chi_{j_2}\|_1 \geq \delta^{0+}\,\|\mu\|_2 \geq \delta^{0+}2^{j_2}\,\|\chi_{j_2}\|_2\,.$$

So, since $2^{j_2}|A_{j_2}| \leq 1$,

$$2^{j_1/2}|A_{j_1}|^{1/2} \geq 2^{j_1-j_2/2}|A_{j_1}|^{1/2} \geq 2^{j_1}|A_{j_1}|^{1/2}|A_{j_2}|^{1/2} \geq \delta^{0+}. \tag{19}$$

Similarly,

$$2^{j_1/2-j_2/2} \geq 2^{j_1/2}|A_{j_2}|^{1/2} \geq \delta^{0+},$$

which implies

$$2^{j_1} < 2^{j_2} \leq \delta^{0-}2^{j_1}.$$

Since $2^{j_2}|A_{j_2}| \leq 1$, using Young's inequality and (17), we thus have

$$\begin{aligned}\delta^{0+}2^{-2j_2}|A_{j_1}| &\leq \langle \chi_{j_1} * \chi_{j_2}, \chi_{j_1} * \chi_{j_2}\rangle \leq \|\chi_{j_2}\|_2\,\|\chi_{j_1} * \chi_{j_1} * \chi_{j_2}\|_2 \\ &\leq \|\chi_{j_2}\|_2\,\|\chi_{j_2}\|_1\,\|\chi_{j_1} * \chi_{j_1}\|_2 \leq 2^{-3j_2/2}\,\|\chi_{j_1} * \chi_{j_1}\|_2\,.\end{aligned}$$

Hence,

$$\|\chi_{j_1} * \chi_{j_1}\|_2^2 \geq \delta^{0+}2^{-j_2}|A_{j_1}|^2 \geq \delta^{0+}2^{-j_1}|A_{j_1}|^2 \geq \delta^{0+}|A_{j_1}|^3. \tag{20}$$

Use a version of Balog–Szemeredi–Gowers theorem proved in [Tao08]. Denote

$$\mathcal{K} = B_r(1) \quad \text{with} \quad r = \delta^{-C_3(\gamma)\varepsilon} = \delta^{0-},$$

a compact subset of $\mathsf{SL}_2(\mathbb{R})$, with $C_3(\gamma) \sim C_2(\gamma)$ to be determined. Specifically, if ε is small enough, then

$$A_{j_1} \subset \mathcal{K}.$$

The *multiplicative energy* of A_{j_1} is $\|\chi_{j_1} * \chi_{j_1}\|_2^2$. Equation (20) implies that A_{j_1} has high energy. Theorem 5.4 (or, more precisely, its proof) in [Tao08] implies that, for the appropriate $C_3(\gamma)$, there exists $H \subset \mathcal{K}$ which is an approximate group, namely,

$$H = H^{-1}$$

and there exists a finite set $Y \subset \mathcal{K}$ of size

$$|Y| \leq \delta^{0-} \tag{21}$$

satisfying

$$HH \subset YH \tag{22}$$

so that

$$\delta^{0+}|A_{j_1}| \leq |H| \leq \delta^{0-}|A_{j_1}|. \tag{23}$$

In addition, there is $y \in \mathcal{K}$ such that

$$|A_1| \geq \delta^{0+}|A_{j_1}|, \tag{24}$$

where

$$A_1 = A_{j_1} \cap yH.$$

Finally, define

$$A = \left((A_1^{-1}A_1) \cup (A_1A_1^{-1})\right) \cap B_\alpha(1),$$

for $\alpha > 0$ as in Theorem 16. Hence,

$$|A| \geq \delta^{0+}|A_1| \geq \delta^{0+}|A_{j_1}|. \tag{25}$$

We now prove that A violates the product theorem. We first show that it violates the conclusion of the product theorem and then show that it satisfies the assumptions of the product theorem.

Using (18) and Young's inequality,

$$2^{j_1+j_2}|A_{j_2}|^{1/2}|A_{j_1}| = 2^{j_1+j_2}\left\|\chi_{j_2}\right\|_2\left\|\chi_{j_1}\right\|_1 \geq \delta^{0+}\left\|\mu\right\|_2 \geq \delta^{0+}2^{j_2}|A_{j_2}|^{1/2}.$$

Hence, using (24),

$$\mu(yH) \geq \mu(A_1) \geq \delta^{0+}2^{j_1}|A_{j_1}| \geq \delta^{0+}. \tag{26}$$

On the other hand,

$$\mu(yH) \lesssim \delta^{-3} \max_{z \in \mathsf{supp}(\nu^{(\ell_0)})} \left|yH \cap B_{\delta^{1-}}(z)\right|.$$

So, there is $z_0 \in \mathcal{K}$ so that

$$|H \cap S| \geq \delta^{3+},$$

with

$$S = B_{\delta^{1-}}(z_0).$$

Let Z be a maximal set of points in H so that for all $z \neq z'$ in Z,

$$zS \cap z'S = \emptyset.$$

Bound,

$$\delta^{0-}|H| \geq |HH| \geq |Z|\,|H \cap S| \geq \delta^{3+}\mathcal{N}_\delta(H).$$

Hence,

$$\mathcal{N}_\delta(H) \leq \delta^{-3-}|H|. \tag{27}$$

Finally,

$$\mathcal{N}_\delta(AAA) \lesssim \mathcal{N}_\delta(H_{(6)}) \leq \delta^{-3-}|H| \leq \delta^{-3-}|A| \leq \delta^{0-}\mathcal{N}_\delta(A).$$

So, indeed, the conclusion of the product theorem does not hold. It remains to prove that A satisfies the assumptions of the product theorem.

First,

$$A = A^{-1}.$$

The second thing we show is that A is not too small or too large. Equation (18) implies

$$\delta^{0+}\|\mu\|_2 \leq 2^{j_1+j_2}\|\chi_{j_1} * \chi_{j_2}\|_2 \leq 2^{j_1}\|\chi_{j_1}\|_2 2^{j_2}\|\chi_{j_2}\|_1 \leq 2^{j_1}|A_{j_1}|^{1/2},$$

which implies

$$\delta^{-\gamma+} \leq 2^{j_1}|A_{j_1}|^{1/2} \lesssim \|\mu\|_2 \leq \delta^{-3/2+\gamma}.$$

Thus,

$$\delta^{-2\gamma+}|A_{j_1}| \leq (2^{j_1}|A_{j_1}|)^2 \leq 1$$

and, using (19),

$$\delta^{0+} \leq (2^{j_1}|A_{j_1}|)^2 \lesssim \delta^{-3+2\gamma}|A_{j_1}|.$$

Therefore,

$$\delta^{3-2\gamma+} \leq |A_{j_1}| \leq \delta^{2\gamma-},$$

which implies, using (23),

$$\delta^{3-2\gamma+} \leq |H| \leq \delta^{2\gamma-}.$$

Therefore, using (21), (22), (25), and (27),

$$\delta^{-2\gamma+} \leq \delta^{-3+}|A_{j_1}| \leq \delta^{-3+}|A| \leq \mathcal{N}_\delta(A) \leq \delta^{-3-}|H| \leq \delta^{-3+2\gamma-},$$

or

$$\mathcal{N}_\delta(A) = \delta^{-3+\sigma_0},$$

with $\sigma_1 < \sigma_0 < 3 - \sigma_1$ and $\sigma_1 = 2\gamma-$.

Thirdly, we prove that A is well-distributed: Let $\varepsilon_5 = \varepsilon_5(\sigma_1, \tau) > 0$ be as given by Theorem 16 for $\tau > 0$ a universal constant to be determined, and let $\delta < \rho < \delta^{\varepsilon_5}$. We prove that there is a finite set $X \subset A$ so that $|X| \geq \rho^{-\tau}$ and for every $x \neq x'$ in X we have $\|x - x'\|_2 \geq \rho$. Equation (26) says $\mu(A_1) \geq \delta^{0+}$. Write $\nu^{(\ell_0)} = \nu^{(\ell)} * \nu^{(\ell_0-\ell)}$, for $\ell < \ell_0$ the largest integer so that

$$Q^{-\ell} > \rho.$$

There thus exists $z_1 \in \mathcal{K}$ so that

$$\nu^{(\ell)}(A_1 z_1) \geq \delta^{0+}.$$

By Lemma 3, for every $x \neq x'$ in $\mathsf{supp}(\nu^{(\ell)}) \subseteq \mathcal{W}_\ell(\mathcal{G})$,

$$\left\|x - x'\right\|_2 \geq Q^{-\ell} > \rho.$$

By Proposition 15,

$$\nu^{(\ell)}(A_1 z_1) \leq |\mathcal{W}_\ell(\mathcal{G}) \cap A_1 z_1| \left(\frac{2|\mathcal{G}| - 1}{|\mathcal{G}|^2}\right)^{\ell/2}.$$

Thus, using Lemma 3 again,

$$\mathcal{N}_\rho(A) \geq \delta^{0+} \mathcal{N}_\rho(A_1 z_1) \geq \delta^{0+} |\mathcal{W}_\ell(\mathcal{G}) \cap A_1 z_1| \geq \delta^{0+} \left(\tfrac{|\mathcal{G}|^2}{2|\mathcal{G}|-1}\right)^{\ell/2} \geq \rho^{-\tau},$$

for $\tau \sim 1$.

It remains to show that A contains matrices with certain properties. That is, w.r.t. every basis in a bounded domain, there is $g \in A_{(4)}$ so that $|g_{1,2} g_{2,1}| \geq \delta^{\varepsilon_5}$. Fix a basis diagonalizing some matrix in $\mathsf{SL}_2(\mathbb{R}) \cap B_1(1)$. Choose ℓ_1 large, to be determined. By Proposition 8 from [BG08], since the elements of $\mathcal{G}$ freely generate a group, if $S \subset \mathcal{W}_{\ell_1}(\mathcal{G})$ is so that for all $g_1, g_2, g_3, g_4 \in S$, the bi-commutator $[[g_1, g_2], [g_3, g_4]]$ is 1, then $|S| \leq \ell_1^6$. As above, there is $z_2 \in \mathcal{K}$ so that

$$|\mathcal{W}_{\ell_1}(\mathcal{G}) \cap A_1 z_2| \geq \delta^{0+} \left(\frac{|\mathcal{G}|^2}{2|\mathcal{G}| - 1}\right)^{\ell_1/2}.$$

The set $A_1 z_2$ is contained in a ball of radius $r' = \delta^{0-}$ around 1. Cover the ball of radius r' around 1 by balls of radius $\beta = \alpha/(r'+1) \geq \delta^{0+}$. There thus exists $z_3 \in \mathcal{W}_{\ell_1}(\mathcal{G}) \cap A_1 z_2$ so that

$$|\mathcal{W}_{\ell_1}(\mathcal{G}) \cap A_1 z_2 \cap B_\beta(z_3)| \geq \delta^{0+} \left(\frac{|\mathcal{G}|^2}{2|\mathcal{G}| - 1}\right)^{\ell_1/2} > \ell_1^6$$

(the last inequality is the first property ℓ_1 should satisfy). Hence, there are

$$g_1, g_2, g_3, g_4 \in (\mathcal{W}_{\ell_1}(\mathcal{G}) \cap A_1 z_2 \cap B_\beta(z_3)) z_3^{-1} \subset A_1 A_1^{-1}$$

with non-trivial bi-commutator. For every $g' \in \{g_1, g_2, g_3, g_4\}$,

$$\left\|g' - 1\right\|_2 \leq \left\|g' z_3 - z_3\right\|_2 (r' + 1) \leq \beta(r' + 1) = \alpha,$$

which implies

$$g' \in A.$$

If $g' \in \{g_1, g_2, g_3, g_4\}$ is so that $|(g')_{1,2}(g')_{2,1}| \neq 0$, then

$$|(g')_{1,2}(g')_{2,1}| \geq Q^{-20\ell_1} \geq \delta^{\varepsilon_5}$$

(this is the second property ℓ_1 should satisfy). In this case, we are done. Otherwise, recall that if four 2×2 matrices are either all upper triangular or all lower triangular,

then they have a trivial bi-commutator. So, w.l.o.g. g_1 is lower triangular and g_2 is upper triangular, which implies that $g_1 g_2$ has the required property. □

7 A Product Theorem

In this section we prove the product theorem, Theorem 16. The proof consists of several parts given in the following sub-sections (the outline of the proof follows [BG07], but the proof in our case is more elaborate). The theorem is finally proved in Section 7.5. We start this section with a brief outline of the proof of the product theorem. We note that not only field properties are used but also metric properties, the argument is a multi-scale one. Here are the steps of the proof (ignoring many technicalities).

We wish to prove that a set A with certain properties becomes larger when multiplied by itself.

(i) Assume toward a contradiction that $A_{(3)}$ is not larger than A.
(ii) Assuming (i), find a set V of commuting matrices which is not too small and is close to $A_{(2)}$. To do so, use the trace map, the pigeon-hole principle and a non-commutative version of the Plünnecke–Ruzsa theorem.
(iii) If V is concentrated in a small ball, then AV will "move V around" and hence AV will be much bigger than A. This is a contradiction, as AV is close to $A_{(3)}$.
(iv) Otherwise, V is not concentrated on any ball, which means that it is well-distributed. In this case, use the discretized ring conjecture, which roughly states that a well-distributed set in $\mathbb{R}$ becomes larger under sums and products. To move from $\mathsf{SL}_2(\mathbb{R})$ to $\mathbb{R}$, use matrix-trace, which translates matrix-product to sums and products in the field.

In fact, the size of V obtained is roughly $|A|^{1/3}$. To get back to the "correct" order of magnitude, we use that A is far from strict subgroups in that it contains a matrix g so that $g_{1,2}g_{2,1}$ is far from zero (w.r.t. any basis change). This property of A is used to show that the size of $VgVgV$ is $|V|^3 \sim |A|$.

7.1 Finding commuting matrices. In this sub-section we show that, under some non-degeneracy conditions, a set of matrices induces a not-too-small set of commuting matrices. To prove this, we also show that a set of matrices induces a not-too-small trace-set. We start by stating the results. The proofs follow.

The *trace* of a matrix g is $\mathsf{Tr} g = g_{1,1} + g_{2,2}$. Every g in $\mathsf{SL}_2(\mathbb{C})$ with $|\mathsf{Tr} g| \neq 2$ can be diagonalized (elements g in $\mathsf{SL}_2(\mathbb{R})$ with $|\mathsf{Tr} g| < 2$ have complex eigenvalues, so we must consider $\mathsf{SL}_2(\mathbb{C})$). Define Diag to be the set of diagonal matrices v in $\mathsf{SL}_2(\mathbb{C})$ so that $\mathsf{Tr} v \in \mathbb{R}$.

The following lemma shows that the trace-set of a set is not too small, at least after multiplying by one of four matrices that are "independent" enough. To numerically capture independence of four matrices $g_0, g_1, g_2, g_3 \in \mathsf{SL}_2(\mathbb{R})$ use the volume they define as real vectors:

$$\det(g_0,g_1,g_2,g_3) := \det \begin{pmatrix} (g_0)_{1,1} & (g_1)_{1,1} & (g_2)_{1,1} & (g_3)_{1,1} \\ (g_0)_{1,2} & (g_1)_{1,2} & (g_2)_{1,2} & (g_3)_{1,2} \\ (g_0)_{2,1} & (g_1)_{2,1} & (g_2)_{2,1} & (g_3)_{2,1} \\ (g_0)_{2,2} & (g_1)_{2,2} & (g_2)_{2,2} & (g_3)_{2,2} \end{pmatrix}.$$

LEMMA 17. *Think of* $\mathsf{SL}_2(\mathbb{R})$ *as a subset of* $\mathbb{R}^4$*, and let* $g_0,g_1,g_2,g_3 \in \mathsf{SL}_2(\mathbb{R}) \cap B_{1/2}(1)$ *be so that*

$$|\det(g_0,g_1,g_2,g_3)| \geq \delta^{0+}, \tag{28}$$

and let $A \subset \mathsf{SL}_2(\mathbb{R}) \cap B_{1/2}(1)$*. Then, there is* $I \subset \{0,1,2,3\}$ *of size* $|I|=3$ *so that*

$$\prod_{i\in I} \mathcal{N}_\delta(\mathsf{Tr} g_i^{-1} A) \geq \delta^{0+} \mathcal{N}_\delta(A).$$

The following lemma allows to find a commuting set of matrices via trace.

LEMMA 18. *Let* $A \subset \mathsf{SL}_2(\mathbb{C}) \cap B_\alpha(1), \alpha > 0$ *a small constant, be so that* $\mathsf{dist}(A, \pm 1) \geq \delta^{0+}$*. Then, there exists a set* $V \subset \mathsf{SL}_2(\mathbb{C})$ *of commuting matrices so that*

$$\mathcal{N}_\delta(V) \geq \delta^{0+} \frac{\mathcal{N}_\delta(\mathsf{Tr} A)\mathcal{N}_\delta(A)}{\mathcal{N}_\delta(A^2A^{-1})},$$

and every $v \in V$ *satisfies* $\mathsf{dist}(v, A^{-1}A) \leq \delta^{1-}$.

We shall also need the following corollary of the two lemmas.

COROLLARY 19. *Let* $A \subset \mathsf{SL}_2(\mathbb{R}) \cap B_\alpha(1), \alpha > 0$ *a small constant. Let* $g_1,g_2,g_3 \in \mathsf{SL}_2(\mathbb{R}) \cap B_\alpha(1)$ *be so that*

$$|\det(1,g_1,g_2,g_3)| \geq \delta^{0+}.$$

Then, there is a set of commuting matrices $V \subset \mathsf{SL}_2(\mathbb{C})$ *so that there is* $g_0 \in \{1,g_1,g_2,g_3\}$ *so that*

$$\mathcal{N}_\delta(V) \geq \delta^{0+} \frac{\mathcal{N}_\delta(A)^{4/3}}{\mathcal{N}_\delta(Ag_0^{-1}AA^{-1})},$$

and every $v \in V$ *satisfies* $\mathsf{dist}(v, A^{-1}A) \leq \delta^{1-}$.

Proof of Lemma 17. For $i \in \{0,1,2,3\}$, denote

$$g_i' = \begin{pmatrix} d_i & -c_i \\ -b_i & a_i \end{pmatrix},$$

where

$$g_i = \begin{pmatrix} a_i & b_i \\ c_i & d_i \end{pmatrix}.$$

By (28),

$$|\det(g_0', g_1', g_2', g_3')| = |\det(g_0, g_1, g_2, g_3)| \geq \delta^{0+}.$$

Hence, let $A' \subset A$ be contained in some ball of radius δ^{0+} so that

$$\mathcal{N}_\delta(A) \leq \delta^{0-}\mathcal{N}_\delta(A'),$$

and so that there is a set $I \subset \{0,1,2,3\}$ of size $|I| = 3$ so that

$$\mathcal{N}_\delta(A') \leq \delta^{0-}\mathcal{N}_\delta(PA'),$$

where P is the projection to the sub-space $\mathsf{span}\{g_i' : i \in I\}$ (the map $g \mapsto Pg$ restricted to a small ball is a diffeomorphism with bounded distortion). For every $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ in $\mathsf{SL}_2(\mathbb{R})$,

$$\mathsf{Tr} g_i^{-1} g = d_i a - b_i c - c_i b + a_i d = \left\langle g, g_i' \right\rangle,$$

with the standard inner product over $\mathbb{R}^4$. Thus,

$$\mathcal{N}_\delta(PA') \leq \delta^{0-} \prod_{i \in I} \mathcal{N}_\delta(\mathsf{Tr} g_i^{-1} A') \leq \delta^{0-} \prod_{i \in I} \mathcal{N}_\delta(\mathsf{Tr} g_i^{-1} A). \qquad \square$$

Proof of Lemma 18. Choose $T \subset \mathsf{Tr}A$ so that

$$|T| \sim \mathcal{N}_\delta(\mathsf{Tr}A), \tag{29}$$

and so that for all $t \neq t'$ in T,

$$|t - t'|, |t-2|, |t+2| > 2\delta.$$

(If $\mathcal{N}_\delta(\mathsf{Tr}A)$ is small, the lemma trivially holds.) Since trace is continuous,

$$\sum_{t \in T} \mathcal{N}_\delta\left(\left\{g \in A^2A^{-1} : |\mathsf{Tr} g - t| < \delta/4\right\}\right) \lesssim \mathcal{N}_\delta(A^2A^{-1}).$$

There thus exists $t_0 \in T$ so that the set

$$A_0 = \{g \in A^2A^{-1} : |\mathsf{Tr} g - t_0| < \delta/4\}$$

satisfies

$$\mathcal{N}_\delta(A_0) \lesssim \frac{\mathcal{N}_\delta(A^2A^{-1})}{|T|}.$$

Choose $g_0 \in A$ so that $\mathsf{Tr} g_0 = t_0$.

Choose $A_1 \subset A_0$ so that

$$|A_1| = \mathcal{N}_\delta(A_0)$$

and

$$A_0 \subset \bigcup_{g \in A_1} B_\delta(g). \tag{30}$$

For $g \in A_1$, define (with a slight abuse of notation)

$$A_g = \{x \in A : x g_0 x^{-1} \in B_\delta(g)\}.$$

Since for every x we have $\mathsf{Tr} x g_0 x^{-1} = \mathsf{Tr} g_0 = t_0$, for every $x \in A$ we have $x g_0 x^{-1} \in A_0$. Equation (30) thus implies

$$A = \bigcup_{g \in A_1} A_g.$$

Hence, there is $g_1 \in A_1$ so that

$$\mathcal{N}_\delta(A_{g_1}) \geq \frac{\mathcal{N}_\delta(A)}{|A_1|} = \frac{\mathcal{N}_\delta(A)}{\mathcal{N}_\delta(A_0)} \gtrsim \frac{\mathcal{N}_\delta(A)}{\mathcal{N}_\delta(A^2 A^{-1})} |T|. \tag{31}$$

Fix $x_1 \in A_{g_1}$. By definition, for every $x \in A_{g_1}$,

$$\left\| x g_1 x^{-1} - x_1 g_1 x_1^{-1} \right\| \leq 2\delta.$$

Since A is bounded,

$$\|y g_1 - g_1 y\| \lesssim \delta,$$

where

$$y = x_1^{-1} x \in x_1^{-1} A_{g_1}.$$

Since $g_1 \in A$ is far from ± 1, conclude that diagonalizing g_1 makes $x_1^{-1} A$ close to diagonal: Since $|\mathsf{Tr} g_1| \neq 2$, there exists a matrix u so that $v_1 = u g_1 u^{-1}$ is diagonal. By assumption on A,

$$\mathsf{dist}(v_1, \pm 1) \sim \mathsf{dist}(g_1, \pm 1) \geq \delta^{0+}.$$

So,

$$|(v_1)_{1,1} - (v_1)_{2,2}| \geq \delta^{0+}.$$

In addition,

$$\left\| u y u^{-1} v_1 - v_1 u y u^{-1} \right\| \lesssim \delta.$$

Hence,

$$|(u y u^{-1})_{1,2}|, |(u y u^{-1})_{2,1}| \lesssim \delta^{1-}.$$

Since $|\det(u y u^{-1})| = 1$, there is thus a diagonal $v \in \mathsf{SL}_2(\mathbb{C})$ so that

$$\left\| u y u^{-1} - v \right\| \lesssim \delta^{1-}.$$

We can thus conclude that $x_1^{-1}A_{g_1} \subset A^{-1}A$ is in a (δ^{1-})-neighborhood of a set $V \subset \mathsf{SL}_2(\mathbb{C})$ of commuting matrices. In particular,

$$\mathcal{N}_\delta(V) \geq \delta^{0+}\mathcal{N}_\delta(A_{g_1}).$$

Equations (29) and (31) imply the claimed lower bound on $\mathcal{N}_\delta(V)$.

Proof of Corollary 19. Since $|\det(1, g_1, g_2, g_3)| \geq \delta^{0+}$, the pairwise distances between $\pm 1, \pm g_1, \pm g_2, \pm g_3$ are at least δ^{0+}. Thus, there exists a subset A' of A so that

$$\mathcal{N}_\delta(A') \geq \delta^{0+}\mathcal{N}_\delta(A)$$

and

$$\mathsf{dist}(A', \{\pm 1, \pm g_1, \pm g_2, \pm g_3\}) \geq \delta^{0+}.$$

By Lemma 17, there exists $g_0 \in \{1, g_1, g_2, g_3\}$ so that

$$\mathcal{N}_\delta(\mathsf{Tr} g_0^{-1}A') \geq \delta^{0+}\mathcal{N}_\delta(A')^{1/3}.$$

Now, apply Lemma 18 on the set $g_0^{-1}A'$ to complete the proof. □

7.2 Trace expansion via discretized ring conjecture. The following lemma is the main result of this section. The lemma roughly tells us that if a set V of commuting matrices is well-distributed then adding a non-commuting element to V makes its trace-set grow under products.

LEMMA 20. *For every $0 < \sigma < 2$ and $0 < \kappa < 1$, there are $\varepsilon_4, \varepsilon_4' > 0$ so that the following holds. Let $V \subset \mathsf{SL}_2(\mathbb{C}) \cap B_\alpha(1), \alpha > 0$ a small constant, be so that $V = V^{-1}$, so that $\mathsf{dist}(v, \mathsf{Diag}) \leq \delta^{1-\varepsilon_4'}$ for all v in V, so that*

$$\mathcal{N}_\delta(V) = \delta^{-\sigma},$$

and so that for all $\delta < \rho < \delta^{\varepsilon_4}$,

$$\max_a \mathcal{N}_\delta(V \cap B_\rho(a)) < \rho^\kappa \delta^{-\sigma}. \tag{32}$$

Let $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathsf{SL}_2(\mathbb{CC}) \cap B_\alpha(1)$ be so that $\mathsf{Tr} g \in \mathbb{R}$ and $|bc| \geq \delta^{\varepsilon_4}$. Then,

$$\mathcal{N}_\delta(\mathsf{Tr} WgWg) \geq \delta^{-\sigma-\varepsilon_4},$$

where $W = V_{(8)}$.

The starting point here is the discretized ring conjecture. This conjecture was first prove in [Bou03] and later strengthened in [BG07], see Proposition 3.2 in [BG07].

LEMMA 21. *For all $0 < \sigma, \kappa < 1$, there is $\varepsilon_2 > 0$ so that for all $\delta > 0$ small, the following holds. Let $A \subset [-1,1]$ be a union of δ-intervals so that*

$$|A| = \delta^{1-\sigma}$$

and for all $\delta < \rho < \delta^{\varepsilon_2}$,

$$\max_a |A \cap B_\rho(a)| < \rho^\kappa |A|.$$

Then,

$$|A + A| + |AA| > \delta^{1-\sigma-\varepsilon_2}.$$

The discretized ring conjecture was used in [BG07] to prove "scalar amplification," i.e., the following proposition.

PROPOSITION 22. *For all $0 < \sigma, \kappa < 1$, there is $\varepsilon_3 > 0$ so that the following holds. Let $S \subset \mathbb{C}$ be a subset of the complex unit circle, so that S is a union of δ-arcs, $\delta > 0$ small enough, so that $S = S^{-1}$, so that*

$$|S| = \delta^{1-\sigma}$$

(size is measured in the unit circle), and so that for all $\delta < \rho < \delta^{\varepsilon_3}$,

$$\max_a |S \cap B_\rho(a)| < \rho^\kappa |S|. \tag{33}$$

If $\gamma, \lambda \in \mathbb{R}$ are so that $\gamma > 0, |\lambda| \geq \delta^{\varepsilon_3}$, then the set

$$D = \{xy + \gamma/(xy) + \lambda(x/y + y/x) : x, y \in S_{(4)}\}$$

satisfies

$$\mathcal{N}_\delta(D) \geq \delta^{-\varepsilon_3 - \sigma}.$$

We also need and prove the following variant of scalar amplification.

PROPOSITION 23. *For all $0 < \sigma, \kappa < 1$, there is $\varepsilon_3 > 0$ so that the following holds. Let $S \subset [1/2, 2]$ be a union of δ-intervals, $\delta > 0$ small enough, so that $S = S^{-1}$, so that*

$$|S| = \delta^{1-\sigma},$$

and so that for all $\delta < \rho < \delta^{\varepsilon_3}$,

$$\max_a |S \cap B_\rho(a)| < \rho^\kappa |S|. \tag{34}$$

If $\gamma, \lambda \in \mathbb{R}$ are so that $\gamma > 0, |\lambda| \geq \delta^{\varepsilon_3}$, then the set

$$D = \{xy + \gamma/(xy) + \lambda(x/y + y/x) : x, y \in S_{(4)}\}$$

satisfies

$$\mathcal{N}_\delta(D) \geq \delta^{-\varepsilon_3 - \sigma}.$$

Lemma 20 follows from scalar amplification.

Proof of Lemma 20. Let $V_0 \subset \mathsf{Diag}$ be so that

$$\mathsf{dist}(v, V_0) \leq \delta_0 = \delta^{1-}$$

for all v in V and $\mathsf{dist}(v_0, V) \leq \delta_0$ for all v_0 in V_0. Specifically, for all $\delta_0 < \rho < {\delta_0}^{2\varepsilon_4}$,

$$\max_a \mathcal{N}_{\delta_0}(V_0 \cap B_\rho(a)) \leq \delta^{0-} \max_a \mathcal{N}_{\delta_0}(V \cap B_\rho(a)) \leq \delta^{0-} \rho^\kappa \delta^{-\sigma}. \tag{35}$$

Observe

$$\mathsf{Tr} \begin{pmatrix} x & 0 \\ 0 & 1/x \end{pmatrix} g \begin{pmatrix} y & 0 \\ 0 & 1/y \end{pmatrix} g = a^2 xy + d^2/(xy) + bc(x/y + y/x). \tag{36}$$

Write

$$V_0 = \left\{ \begin{pmatrix} x & 0 \\ 0 & 1/x \end{pmatrix} : x \in T \right\}.$$

The set T is contained in the real numbers union the complex unit circle. Denote by $T_1 = T \cap \mathbb{R}$, and $T_2 = T \setminus T_1$. First, assume

$$\mathcal{N}_{\delta_0}(T_1) \sim \mathcal{N}_{\delta_0}(V_0). \tag{37}$$

Define S_1 to be a δ_0-neighborhood of T_1. Thus,

$$|S_1| = \delta_0^{1-\sigma_1}$$

with $\sigma_1 \geq \sigma/2$. Equation (35) implies that S_1 satisfies (34) with $\kappa_1 = \kappa/2$. As in Proposition 23, denote

$$D_1 = a^2\{xy + \gamma/(xy) + \lambda(x/y + y/x) : x, y \in (S_1)_{(4)}\}.$$

with $\gamma = (d/a)^2$ and $\lambda = bc/a^2$. Observe, $ad - bc = 1$ and $a + d \in \mathbb{R}$ imply $d/a \in \mathbb{R}$ and $bc/a^2 \in \mathbb{R}$. In addition, $|\lambda| \geq \delta_0^{0+}$. The proposition thus implies

$$\mathcal{N}_{\delta_0}(D_1) \geq \delta_0^{-\varepsilon_3 - 1} |S_1| \geq \delta^{-\varepsilon_3 - \sigma +}.$$

Using (36), conclude

$$\mathcal{N}_\delta(\mathsf{Tr} W g W g) \geq \delta^{-\sigma - \varepsilon_3 +}.$$

When (37) does not hold, consider T_2 and use Proposition 22 instead of Proposition 23. □

Proof of Proposition 23. Assume towards a contradiction that the proposition does not hold. W.l.o.g., for every s in S,

$$\mathsf{dist}(s, \{\gamma^{1/4}, 1\}) \geq \delta^{0+}. \tag{38}$$

We first find a set A so that $A+A$ is not much larger than A. If $s, s' \in S$, then $x = s'/s \in S_{(2)}$ and $y = ss' \in S_{(2)}$ satisfy $xy = s'^2$ and $y/x = s^2$. By assumption, we can thus conclude

$$\left|\left\{(s'^2 + \gamma/s'^2) + \lambda(s^2 + 1/s^2) : s', s \in S_{(2)}\right\}\right| \lesssim \delta^{-\varepsilon_3}|S|.$$

Denote

$$A = \{\lambda(s^2 + 1/s^2) : s \in S_{(2)}\}$$

and

$$A' = \{s'^2 + \gamma/s'^2 : s' \in S_{(2)}\}.$$

Since $|\lambda| \geq \delta^{0+}$,

$$|A| \geq \delta^{0+}|S|.$$

By (38), the derivative of the map $s' \mapsto s'^2 + \gamma/s'^2$ is bounded away from zero in the relevant range. Thus,

$$|A'| \geq \delta^{0+}|S|.$$

Ruzsa's inequality in measure version for open sets $A, A' \subset \mathbb{R}$ states $|A+A| \leq |A+A'|^2/|A'|$ (see, e.g., Lemma 3.2 in [Tao08]). Therefore,

$$|A+A| \leq \delta^{0-}|S|. \tag{39}$$

We now find a set that does not significantly increase its size under sums and products. Define

$$A_1 = \{s^2 + 1/s^2 : s \in S\}.$$

By (38),

$$|A_1| \geq \delta^{0+}|S|.$$

Hence, by (39), since $|\lambda| \geq \delta^{0+}$,

$$|A_1 + A_1| \leq \delta^{0-}|A+A| \leq \delta^{0-}|A_1|.$$

Observe

$$(s_1^2 + 1/s_1^2)(s_2^2 + 1/s_2^2) = ((s_1 s_2)^2 + 1/(s_1 s_2)^2) + ((s_1/s_2)^2 + 1/(s_1/s_2)^2).$$

Hence, using (39), since $|\lambda| \geq \delta^{0+}$,

$$|A_1 A_1| \leq \delta^{0-}|A+A| \leq \delta^{0-}|A_1|.$$

So,

$$|A_1 + A_1| + |A_1 A_1| \leq \delta^{0-}|A_1|.$$

If $\varepsilon_3 > 0$ is small enough, we can set $0 < \sigma' < 1$ so that

$$|A_1| = \delta^{1-\sigma'}.$$

Choose $\kappa' = \kappa/2$. Set $\varepsilon_2 = \varepsilon_2(\sigma', \kappa') > 0$ as in Lemma 21. If $\varepsilon_3 > 0$ is small enough, then for every $\delta < \rho < \delta^{\varepsilon_2}$,

$$\max_a |A_1 \cap B_\rho(a)| \leq \delta^{0-} \max_a |S \cap B_\rho(a)| < \delta^{0-}\rho^\kappa |S| \leq \delta^{0-}\rho^\kappa |A_1| \leq \rho^{\kappa'}|A_1|.$$

This contradicts Lemma 21. □

7.3 Expansion using a non-commuting element. We shall use the following variant of a lemma from [BG07], see [H08] as well. Roughly, the lemma states that adding a non-commuting element to a commuting set of matrices makes it grow under products.

LEMMA 24. *Let $V \subset \mathsf{SL}_2(\mathbb{C}) \cap B_\alpha(1), \alpha$ a small constant, be so that $\mathsf{dist}(v, \mathsf{Diag}) \leq \delta^{1-}$ for all v in V. Let $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathsf{Diag} \cap B_\alpha(1)$ be so that $|bc| \geq \delta^{0+}$. Then,*

$$\mathcal{N}_\delta(VgVgV) \geq \delta^{0+}\mathcal{N}_\delta(V)^3.$$

Proof. Assume

$$\mathcal{N}_\delta(V) > \delta^{0-} \tag{40}$$

(otherwise, the lemma trivially holds). There are several cases to consider.

1. Denote by $\mathsf{Diag}_\mathbb{R}$ the set of matrices in Diag with entries in $\mathbb{R}$. Consider the case that there is a subset of $\mathsf{Diag}_\mathbb{R}$ with comparable metric entropy to that of V: assume that there is $Z \subset \mathbb{R}$ so that $|Z| \geq \delta^{0+}\mathcal{N}_\delta(V)$, so that for all $z \in Z$,

$$\mathsf{dist}\left(\begin{pmatrix} z & 0 \\ 0 & 1/z \end{pmatrix}, V\right) \leq \delta^{1-},$$

and so that for all $z \neq z'$ in Z,

$$|z - z'| > \delta.$$

W.l.o.g., assume that $z \geq \sqrt{d/a}$ (the proof in the other case is similar). Furthermore, by (40), we can assume w.l.o.g. that

$$z - \sqrt{d/a}, |z - 1| \geq \delta^{0+}.$$

For $z = (z_1, z_2, z_3)$ in Z^3, denote

$$M_z = \begin{pmatrix} z_1 & 0 \\ 0 & 1/z_1 \end{pmatrix} g \begin{pmatrix} z_2 & 0 \\ 0 & 1/z_2 \end{pmatrix} g \begin{pmatrix} z_3 & 0 \\ 0 & 1/z_3 \end{pmatrix}.$$

To prove the lemma, we will show that for all $z \neq z'$ in Z^3,

$$\|M_z - M_{z'}\| \geq \delta^{1+}.$$

Observe

$$M_z = \begin{pmatrix} z_1 z_3 (a^2 z_2 + bc/z_2) & (z_1/z_3) b (a z_2 + d/z_2) \\ (z_3/z_1) c (a z_2 + d/z_2) & (1/z_1 z_3)(bc z_2 + d^2/z_2) \end{pmatrix}.$$

Consider the following two cases.

1.1. The first case is when $z_2 > z_2'$. We have two sub-cases to consider.

1.1.1. The first sub-case is $|z_1/z_3 - z_1'/z_3'| \geq \delta^{1+}$. Bound

$$\left|(M_z)_{1,2}\,/(M_z)_{2,1} - (M_{z'})_{1,2}/(M_{z'})_{2,1}\right| = |b/c| \cdot \left|(z_1/z_3)^2 - (z_1'/z_3')^2\right| \geq \delta^{1+}.$$

Thus,

$$\begin{aligned} \delta^{1+} &\leq \left|(M_z)_{1,2}(M_{z'})_{2,1} - (M_{z'})_{1,2}(M_z)_{2,1}\right| \\ &= \left|\big((M_z)_{1,2} - (M_{z'})_{1,2}\big)(M_{z'})_{2,1} + (M_{z'})_{1,2}\big((M_{z'})_{2,1} - (M_z)_{2,1}\big)\right|. \end{aligned}$$

So,

$$\|M_z - M_{z'}\| \geq \delta^{1+}.$$

1.1.2. The second sub-case is $|z_1/z_3 - z_1'/z_3'| < \delta^{1+}$. Bound

$$\begin{aligned} |(M_z)_{1,2} - (M_{z'})_{1,2}| &= |ba|\left|(z_1/z_3)(z_2 + (d/a)/z_2) - (z_1'/z_3')(z_2' + (d/a)/z_2')\right| \\ &\gtrsim |ba|\left|z_2 + (d/a)/z_2 - z_2' + (d/a)/z_2'\right| - \delta^{1+}. \end{aligned}$$

The map $z_2 \mapsto z_2 + (d/a)/z_2$ has derivative at least δ^{0+} for $z_2 \geq \sqrt{d/a} + \delta^{0+}$. So,

$$|(M_z)_{1,2} - (M_{z'})_{1,2}| \geq \delta^{1+}.$$

1.2. The second case is $z_2 = z_2'$ and $(z_1, z_3) \neq (z_1', z_3')$. Assume w.l.o.g. $z_1 \neq z_1'$ (the argument in the other case is similar). Since the entries of $g \begin{pmatrix} z_2 & 0 \\ 0 & 1/z_2 \end{pmatrix} g$ are bounded away from 0 and V is close to 1,

$$\|M_z - M_{z'}\| \geq \delta^{0+} \left\|(z_1 z_3 - z_1' z_3', z_1 z_3' - z_1' z_3)\right\|.$$

Since $\|(z_3, z_3')\| \gtrsim 1$ and $\left|\det \begin{pmatrix} z_1 & -z_1' \\ -z_1' & z_1 \end{pmatrix}\right| \gtrsim \delta$,

$$\left\|(z_1 z_3 - z_1' z_3', z_1 z_3' - z_1' z_3)\right\| \gtrsim \delta.$$

2. Otherwise, there is a subset of $\mathsf{Diag} \setminus \mathsf{Diag}_{\mathbb{R}}$ with comparable metric entropy to that of V: There is a subset of the complex unit circle Z so that $|Z| \geq \delta^{0+} \mathcal{N}_\delta(V)$, so that for all $z \in Z$,

$$\mathsf{dist}\left(\begin{pmatrix} z & 0 \\ 0 & 1/z \end{pmatrix}, V\right) \leq \delta^{1-},$$

and so that for all $z \neq z'$ in Z,

$$|z - z'| > \delta.$$

Assume w.l.o.g. that $\mathsf{dist}(Z,1) \geq \delta^{0+}$. Also assume w.l.o.g. that every element of Z has positive imaginary part (the other case is similar).

2.1. When $z_2 \neq z_2'$, bound

$$\big||(M_z)_{1,2}| - |(M_{z'})_{1,2}|\big| = |ba|\big||z_2 + (d/a)/z_2| - |z_2' + (d/a)/z_2'|\big|.$$

If we denote, $z_2 = e^{i\theta_2}$ and $z_2' = e^{i\theta_2'}$, then

$$\big||z_2+(d/a)/z_2|^2-|z_2' + (d/a)/z_2'|^2\big| = 2(d/a)\big|\cos(2\theta_2)-\cos(2\theta_2')\big| \geq \delta^{0+}|z_2 - z_2'| > \delta^{1+}.$$

Hence,

$$\|M_z - M_{z'}\| \geq \delta^{1+}.$$

2.2. When $z_2 = z_2'$, the argument is similar to the one in case 1.2. above. □

7.4 Finding "independent directions". Roughly, we now show that two non-commuting matrices induce four "independent directions."

Claim 25. *Let $g_1 \in \mathsf{SL}_2(\mathbb{C}) \cap B_1(1)$ be so that $\mathsf{dist}(g_1, \pm 1) \geq \delta^{0+}$ and $\mathsf{Tr} g_1 \neq 2$. Let $g_2 \in \mathsf{SL}_2(\mathbb{C})$ be so that w.r.t. the basis that makes g_1 diagonal $|(g_2)_{1,2}(g_2)_{2,1}| \geq \delta^{0+}$. Then,*

$$|\det(1, g_1, g_2, g_1 g_2)| \geq \delta^{0+}.$$

Proof. Choose a basis so that g_1 is diagonal (this is a linear transformation on the g_i's with bounded away from zero determinant). Denote $\lambda = (g_1)_{1,1}$. In the new basis,

$$|\det(1, g_1, g_2, g_1 g_2)| = |(\lambda - 1/\lambda)((g_1 g_2)_{1,2}(g_2)_{2,1} - (g_1 g_2)_{2,1}(g_2)_{1,2})|.$$

By choice,

$$|\lambda - 1/\lambda| \geq \delta^{0+}.$$

and

$$|(g_2)_{1,2}(g_2)_{2,1}| \geq \delta^{0+}.$$

Hence,

$$|((g_1 g_2)_{1,2}\ (g_2)_{2,1} - (g_1 g_2)_{2,1}(g_2)_{1,2})| = |(\lambda - 1/\lambda)(g_2)_{1,2}(g_2)_{2,1}| \geq \delta^{0+}.$$ □

7.5 Proof of product theorem. *Proof of Theorem 16* Assume towards a contradiction that

$$\mathcal{N}_\delta(AAA) \leq \delta^{0-}\mathcal{N}_\delta(A).$$

By [Tao08], for every finite k,

$$\mathcal{N}_\delta(A_{(k)}) \leq \delta^{0-}\mathcal{N}_\delta(A) \tag{41}$$

as well.

The first step is to find a large, commuting set of matrices. By assumption on A and using Claim 25, choose g_1, g_2, g_3 in $A_{(8)}$ with $|\det(1, g_1, g_2, g_3)| \geq \delta^{0+}$. Equation (41) and Corollary 19 imply that there is a set of commuting matrices $V \subset \mathsf{SL}_2(\mathbb{C})$ so that

$$\mathcal{N}_\delta(V) \geq \delta^{0+}\mathcal{N}_\delta(A)^{1/3} = \delta^{-1+\sigma_0/3+} \tag{42}$$

and so that

$$V \subset \Gamma_{\delta^{1-}}(A_{(2)}).$$

Assume (by perhaps allowing $V \subset \Gamma_{\delta^{1-}}(A_{(4)})$) that $V = V^{-1}$ and

$$V \subset B_{\delta^{3\varepsilon_5}}(1). \tag{43}$$

Proceed according to two cases.

The first case is when V is well-spread, i.e., the conditions for using the discretized ring conjecture are held. Define

$$\sigma = 1 - \sigma_0/3 - \quad \text{and} \quad \kappa = \tau/6$$

so that $\mathcal{N}_\delta(V) = \delta^{-\sigma}$. Assume that for all $\delta < \rho < \delta^{\varepsilon_4}$ with $\varepsilon_4 = \varepsilon_4(\sigma, \kappa)$ from Lemma 20,

$$\max_a \mathcal{N}_\delta(V \cap B_\rho(a)) < \rho^\kappa \delta^{-\sigma}.$$

By assumption on A, there is $g_0 \in A_{(4)}$ so that (w.r.t. the basis that makes V diagonal) the distance between g_0 and 1 is at most a small constant, and $|(g_0)_{1,2}(g_0)_{2,1}| \geq \delta^{\varepsilon_5}$. Even after the basis change $\mathsf{Tr} g_0 \in \mathbb{R}$. Thus, Lemma 20 implies

$$\mathcal{N}_\delta(\mathsf{Tr} W_0) \geq \delta^{-\sigma-\varepsilon_4},$$

where

$$W_0 = W g_0 W g_0 W \quad \text{and} \quad W = V_{(8)}.$$

(Here and below $C > 0$ will be a large universal constant, that may change its value.) By choice,

$$\mathsf{dist}(g_0^2, \pm 1) \gtrsim \delta^{\varepsilon_5}.$$

Thus, using (43),

$$\mathsf{dist}(W_0, \pm 1) \gtrsim \delta^{2\varepsilon_5}.$$

We can hence apply Lemma 18 with W_0 to obtain a set

$$W_1 \subset \Gamma_{\delta^{1-}}(W_0^{-1} W_0)$$

of commuting matrices so that

$$\mathcal{N}_\delta(W_1) \geq \delta^{0+} \frac{\mathcal{N}_\delta(\mathsf{Tr} W_0)\mathcal{N}_\delta(W_0)}{\mathcal{N}_\delta(W_0^2 W_0^{-1})} \geq \delta^{0+} \frac{\delta^{-\sigma-\varepsilon_4}\mathcal{N}_\delta(V g_0 V g_0 V)}{\mathcal{N}_\delta(W_0^2 W_0^{-1})}.$$

By (41) and Lemma 24, we thus have

$$\mathcal{N}_\delta(W_1) \geq \delta^{0+} \frac{\delta^{-\sigma-\varepsilon_4}\mathcal{N}_\delta(V)^3}{\mathcal{N}_\delta(A)}.$$

So, by (42),

$$\mathcal{N}_\delta(W_1) \geq \delta^{-\sigma-\varepsilon_4/2}.$$

Again, we can find $g_1 \in A_{(4)}$ so that (w.r.t. the basis that makes W_1 diagonal) $\mathsf{dist}(g_1, 1)$ is at most a small constant, $\mathsf{Tr} g_1 \in \mathbb{R}$, and $|(g_1)_{1,2}(g_1)_{2,1}| \geq \delta^{0+}$. So, we can apply Lemma 24 again and get

$$\begin{aligned} \mathcal{N}_\delta(A) &\geq \delta^{0+}\mathcal{N}_\delta(W_1 g_1 W_1 g_1 W_1) \geq \delta^{0+}\mathcal{N}_\delta(W_1)^3 \\ &\geq \delta^{-3\sigma-\varepsilon_4/2} = \delta^{-3+\sigma_0-\varepsilon_4/2} = \delta^{-\varepsilon_4/2}\mathcal{N}_\delta(A). \end{aligned}$$

This contradicts (41), and the proof is complete in this case.

The proof in the second case, when V is not well-spread, is simpler. Indeed, we have

$$\mathcal{N}_\delta(V_0) \geq \rho^\kappa \delta^{-\sigma}$$

with

$$V_0 = V \cap B_\rho(a)$$

(reusing notation). So, by Lemma 24,

$$\mathcal{N}_\delta(V_1) \geq \delta^{0+}\mathcal{N}_\delta(V_0)^3 \geq \rho^{3\kappa}\delta^{-3\sigma+},$$

where

$$V_1 = V_0 g_0 V_0 g_0 V_0 \subset \Gamma_{\delta^{1-}}(A_{(C)})$$

with g_0 from above. By assumption on A, there is a finite $X \subset A$ so that

$$|X| \geq \rho^{-\tau}$$

and for all $x \neq x'$ in X, $\|x - x'\| \geq C\rho$. Denote

$$V_2 = \bigcup_{x \in X} xV_1.$$

Therefore,

$$\mathcal{N}_\delta(V_2) \geq |X|\mathcal{N}_\delta(V_1) \geq \rho^{-\tau}\rho^{3\kappa}\delta^{-3\sigma+} \geq \rho^{-\tau/2}\delta^{-3+\sigma_0+} \geq \delta^{-3+\sigma_0-\varepsilon_4\tau/3} = \delta^{0-}\mathcal{N}_\delta(A).$$

Since $V_2 \subset \Gamma_{\delta^{1-}}(A_{(C)})$, we obtained a contradiction to (41), and the proof is complete. □

References

[Bou03] J. Bourgain. On the Erdos–Volkmann and Katz–Tao ring conjectures. *Geometric and Functional Analysis*, 13 (2003), 334–365.

[Bou09] J. Bourgain. Expanders and dimensional expansion. *Comptes Rendus Mathematique*, (7)347 (2009), 357–362.

[Bou12] J. Bourgain. On the Furstenberg measure and density of states for the Anderson–Bernoulli model at small disorder. *Journal d'Analyse Mathématique* (2012) (see arXiv:1101.2148, to appear).

[Bou12] J. Bourgain. Finitely supported measures on $\mathsf{SL}_2(\mathbb{R})$ which are absolutely continuous at infinity. *GAFA Seminar Notes* (2012, to appear).

[BG07] J. Bourgain and A. Gamburd. On the spectral gap for finitely-generated subgroups of SU(2). *Inventiones Mathematicae*, (1)171 (2007), 83–121.

[BG08] J. Bourgain and A. Gamburd. Uniform expansion bounds for Cayley graphs of $\mathrm{SL}_2(\mathrm{F}_p)$. *Annals of Mathematics*, (2)167 (2008), 625–642.

[Bre08] E. Breuillard. A strong Tits alternative (2008, preprint).

[DN06] C.M. Dawson and M.A. Nielsen. The Solovay–Kitaev algorithm. *Quantum Information and Computation*, 6 (2006), 81–95.

[DS08] Z. Dvir and A. Shpilka. Towards dimension expanders over finite fields. In: *Proceedings of the IEEE 23rd Annual Conference on Computational Complexity* (2008), pp. 304–310.

[DW10] Z. Dvir and A. Wigderson. Monotone expanders: constructions and applications. *Theory of Computing*, (1)6 (2010), 291–308.

[GG79] O. Gabber and Z. Galil. Explicit constructions of linear size superconcentrators. In: *Proceedings of 20th Annual Symposium on the Foundations of Computer Science* (1979), pp. 364–370.

[Gow08] W.T. Gowers. Quasirandom groups. *Combinatorics, Probability and Computing*, 17 (2008), 363–387.

[GLW08] V. Guruswami, J. Lee, and A. Wigderson. Euclidean sections of l_1^N with sublinear randomness and error-correction over the reals. *Approx Random* (2008), 444–454.

[H08] H.A. Helfgott. Growth and generation in $\mathrm{SL}_2(\mathrm{Z}/\mathrm{pZ})$. *Annals of Mathematics*, (2)167 (2008), 601–623.

[HLW06] S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(2006), 439–561.

[KL10] V.A. Kaimanovich and V. Le Prince. Matrix random products with singular harmonic measure. *Geometriae Dedicata* (1)150 (2010), 257–279.

[Kes59] H. Kesten. Symmetric random walks on groups. *Transactions of the American Mathematical Society*, 92 (1959), 336–354.

[Kla84] M.M. Klawe. Limitations on explicit constructions of expanding graphs. *SIAM Journal on Computing*, (1)13 (1984), 156–166.

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

[LZ08] A. Lubotzky and Y. Zelmanov. Dimension expanders. *Journal of Algebra*, (2)319 (2008), 730–738.

[RVW02] O. Reingold, S. Vadhan, and A. Wigderson. Entropy waves, the zig-zag graph product, and new constant-degree expanders. *Annals of Mathematics*, (1)155 (2002), 157–187.

[SX91] Sarnak and X. Xue. Bounds for multiplicities of automorphic representations. *Duke Mathematical Journal*, 64 (1991), 207–227.

[SSU01a] K. Simon, B. Solomyak, and M. Urbanski. Hausdorff dimension of limit sets for parabolic IFS with overlap. *Pacific Journal of Mathematics*, (2)201 (2001), 441–478.

[SSU01b] K. Simon, B. Solomyak, and M. Urbanski. Invariant measures for parabolic IFS with overlaps and random continued fractions. *Transactions of the American Mathematical Society*, (12)353 (2001), 5145–5164.

[SS96] M. Sipser and D.A. Spielman. Expander codes. *IEEE Transactions on Information Theory* (6, part 1)42 (1996), 1710–1722

[Tao08] T. Tao. Product set estimates for non-commutative groups. *Combinatorica*, (5)28 (2008), 547–594.

[Ung51] P. Ungar. A theorem on planar graphs. *Journal of the London Mathematical Society*, (4)1 (1951), 256–262.

Jean Bourgain, Institute for Advanced Study, 1 Einstein Drive, Princeton, NJ 08540, USA
bourgain@math.ias.edu

Amir Yehudayoff, Department of Mathematics, Technion–IIT, Haifa 32000, Israel
amir.yehudayoff@gmail.com

Received: February 9, 2012
Revised: July 9, 2012
Accepted: July 23, 2012