LINEAR RANDOM WALKS ON THE TORUS

WEIKUN HE AND NICOLAS DE SAXCÉ

Abstract. We prove a quantitative equidistribution result for linear random walks on the torus, similar to a theorem of Bourgain, Furman, Lindenstrauss and Mozes, but without any proximality assumption. An application is given to expansion in simple groups, modulo arbitrary integers.

1. Introduction

The goal of the present paper is to study the equidistribution of linear random walks on the torus. We are given a probability measure μ on the group $\mathrm{SL}_d(\mathbb{Z})$ of integer matrices with determinant one, and consider the associated random walk $(x_n)_{n\geq 0}$ on the torus $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$, starting from a point x_0 in $\mathbb{T}^d$, and moving at step n following a random element g_n with law μ:

$$x_n = g_n x_{n-1} = g_n \dots g_1 x_0.$$

We say that the measure μ on $\mathrm{SL}_d(\mathbb{Z})$ has some finite exponential moment if there exists $\varepsilon > 0$ such that

$$\int \|g\|^{\varepsilon} \,\mathrm{d}\mu(g) < \infty,$$

where $\|\ \|$ denotes an arbitrary norm on $\mathcal{M}_d(\mathbb{R})$, the space of $d \times d$ matrices with real coefficients. Our goal is to prove the following theorem.

Theorem 1.1 (Equidistribution on the torus)**.** *Let $d \geq 2$. Let μ be a probability measure on $\mathrm{SL}_d(\mathbb{Z})$. Denote by Γ the subsemigroup generated by μ, and by $\mathbb{G} < \mathrm{SL}_d$ the Zariski closure of Γ. Assume that:*

(a) The measure μ has a finite exponential moment;
(b) The only subspaces of $\mathbb{R}^d$ preserved by Γ are $\{0\}$ and $\mathbb{R}^d$;
(c) The algebraic group $\mathbb{G}$ is Zariski connected.

Then, for every irrational point x in $\mathbb{T}^d$, the sequence of measures $(\mu^{*n} * \delta_x)_{n\geq 1}$ converges to the Haar measure in the weak-$*$ topology.

With an additional proximality assumption, this theorem was proved a decade ago by Bourgain, Furman, Lindenstrauss and Mozes [11], and we follow their approach to this problem, via a study of the Fourier coefficients of the law at time n of the random walk on $\mathbb{T}^d$. One advantage of this method – besides being the only one available at the present – is that it yields a quantitative statement, giving a speed of convergence of the random walk, in terms of the diophantine properties of the starting point x, see [11, Theorem A].

Theorem 1.2 (Quantitative equidistribution on the torus)**.** *Under the assumptions of Theorem 1.1, let λ_1 denote the top Lyapunov exponent associated to μ.*

2010 *Mathematics Subject Classification.* Primary .
Key words and phrases. Sum-product, Random walk, Toral automorphism.
W.H. is supported by ERC grant ErgComNum 682150.

Given $\lambda \in (0, \lambda_1)$, there exists a constant $C = C(\mu, \lambda) > 0$ such that for every $x \in \mathbb{T}^d$ and every $t \in (0, 1/2)$, if for some $a \in \mathbb{Z}^d \setminus \{0\}$,

$$|\widehat{\mu^{*n} * \delta_x}(a)| \geq t \quad \text{and} \quad n \geq C \log \frac{\|a\|}{t},$$

then there exists $q \in \mathbb{Z}_{>0}$ and $x' \in (\frac{1}{q}\mathbb{Z}^d)/\mathbb{Z}^d$ such that

$$q \leq \Big(\frac{\|a\|}{t}\Big)^C \quad \text{and} \quad d(x, x') \leq e^{-\lambda n}.$$

One of our motivations for removing any proximality assumption from this theorem was to generalize a theorem of Bourgain and Varjú on expansion in $\mathrm{SL}_d(\mathbb{Z}/qZ)$, where q is an arbitrary integer, to more general simple $\mathbb{Q}$-groups. We briefly describe this application at the end of the paper, in §6.1.

In [11], the proximality assumption is used at several important places, especially in the study of the large scale structure of Fourier coefficients [11, Phase I]. Let us mention the main ingredients we had to bring into our proof in order to overcome this issue.

One important tool in the proof of Bourgain, Furman, Lindenstrauss and Mozes [11] is a discretized projection theorem, due to Bourgain [10, Theorem 5], giving information on the size of the projections of a set $A \subset \mathbb{R}^d$ to lines. But, when the random walk is not proximal, one should no longer project the set to lines, but to subspaces whose dimension equals the proximality dimension of Γ. One approach, of course, would be to generalize Bourgain's theorem to higher dimensions, and this has been worked out by the first author [26]. But it turns out that the natural generalization of Bourgain's theorem, used with the general strategy of [11], only allows to deal with some special cases [28]. Here we take a different route. Instead of working in the space $\mathbb{Z}^d$ of Fourier coefficients, we place ourselves in the simple algebra $E \subset \mathcal{M}_d(\mathbb{R})$ generated by the random walk. This allows us to use the results of the first author on the discretized sum-product phenomenon in simple algebras [25]. Thus, instead of a projection theorem, we use a result on the Fourier decay of multiplicative convolutions in simple algebras, derived in Section 2, and generalizing a theorem of Bourgain for the field of real numbers [10, Theorem 6].

Then, in order to be able to apply this Fourier estimate to the law at time n of the random walk, we have to check some non-concentration conditions. For that, we use a result of Salehi Golsefidy and Varjú [38] on expansion modulo prime numbers in semisimple groups, combined with a rescaling argument, proved with the theory of random walks on reductive groups. In the end, we obtain some Fourier decay theorem for the law at time n of a random walk on $\mathrm{SL}_d(\mathbb{Z})$, Theorem 3.19, which, we believe, bears its own interest and, we hope, will have other applications.

The rest of the proof, corresponding to [11, Phase II], follows more closely the strategy of [11]. But since at several points we had to find an alternative proof to avoid the use of the proximality assumption, we chose to include the whole argument, rather than refer the reader to [11]. We hope that this will make the proof easier to follow.

2. Sum-product, L^2-flattening and Fourier decay

The main objective of this section is to prove that in a simple real algebra, multiplicative convolutions of non-concentrated measures admit a polynomial Fourier decay. The precise statement is given in Theorem 2.1 below.

From now on, E will denote a finite-dimensional real associative simple algebra, endowed with a norm $\|\ \|$. Given a finite Borel measure μ on E and an integer

$s \geq 1$, we write

$$\mu^{*s} = \underbrace{\mu * \cdots * \mu}_{s \text{ times}}$$

for the s-fold multiplicative convolution of μ with itself. In order to ensure the Fourier decay of some multiplicative convolution of the measure μ, we need two assumptions: First, μ should not be concentrated around a linear subspace of E, and second, μ should not give mass to elements of E that are too singular.

To make these requirements more precise, let us set up some notation. For $\rho > 0$ and $x \in E$, let $B_E(x, \rho)$ denote the closed ball in E of radius ρ and centered at x. For a subset $W \subset E$, let $W^{(\rho)}$ denote the ρ-neighborhood of W,

$$W^{(\rho)} = W + B_E(0, \rho).$$

For $a \in E$ define $\det_E(a)$ to be the determinant of the endomorphism $E \to E$, $x \mapsto ax$. Note that since E is simple this quantity is equal to the determinant of $E \to E$, $x \mapsto xa$. For $\rho > 0$, define $S_E(\rho)$, the set of badly invertible elements of E, as

$$S_E(\rho) = \{x \in E \mid |\det_E(x)| \leq \rho\}.$$

Theorem 2.1 (Fourier decay of multiplicative convolutions)**.** *Let E be a normed simple algebra over $\mathbb{R}$ of finite dimension. Given $\kappa > 0$, there exists $s = s(E, \kappa) \in \mathbb{N}$ and $\varepsilon = \varepsilon(E, \kappa) > 0$ such that for any parameter $\tau \in (0, \varepsilon\kappa)$ the following holds for any scale $\delta > 0$ sufficiently small. Let μ be a Borel probability measure on E. Assume that*

(i) $\mu\big(E \setminus B_E(0, \delta^{-\varepsilon})\big) \leq \delta^{\tau}$;
(ii) for every $x \in E$, $\mu(x + S_E(\delta^{\varepsilon})) \leq \delta^{\tau}$;
(iii) for every $\rho \geq \delta$ and every proper affine subspace $W \subset E$, $\mu(W^{(\rho)}) \leq \delta^{-\varepsilon}\rho^{\kappa}$.

Then for all $\xi \in E^*$ with $\|\xi\| = \delta^{-1}$,

$$|\widehat{\mu^{*s}}(\xi)| \leq \delta^{\varepsilon\tau}.$$

For $E = \mathbb{R}$, this is due to Bourgain [10, Lemma 8.43]. Li proved in [33] a similar statement for the semisimple algebra $\mathbb{R} \oplus \cdots \oplus \mathbb{R}$. While a more general statement should hold for any semisimple algebra, we do not pursue in this direction and focus in the present paper only on simple algebras.

2.1. **L^2-flattening.** The aim of this subsection is to prove a sum-product L^2-flattening lemma for simple algebras.

We shall consider both additive and multiplicative convolutions between measures or functions on E. To avoid confusion, we shall use the usual symbol $*$ to denote multiplicative convolution and the symbol $\boxplus$ to denote additive convolution. In the same fashion, for finite Borel measures μ and ν on E, we define $\mu \boxminus \nu$ to be the push forward measure of $\mu \otimes \nu$ by the map $(x, y) \mapsto x - y$.

For a Borel set $A \subset E$, denote by $|A|$ the Lebesgue measure of A. For $\delta > 0$, define $P_\delta = |B_E(0, \delta)|^{-1} \mathbb{1}_{B_E(0,\delta)}$. For absolutely continuous measures such as $\mu \boxplus P_\delta$, by abuse of notation, we write $\mu \boxplus P_\delta$ to denote both the measure and the Radon-Nikodym derivative. For $x \in E$, we write $\mathcal{D}_x$ to denote the Dirac measure at the point x. For $K > 1$, define the set of well invertible elements of E as

$$G_E(K) = \{x \in E^{\times} \mid \|x\|, \|x^{-1}\| \leq K\}.$$

Note that if $x \in G_E(K)$, the left, or right, multiplication by x as a map from E to itself is $O(K)$-bi-Lipschitz.

Proposition 2.2 (L^2-flattening)**.** *Let E be a normed finite-dimensional simple algebra over $\mathbb{R}$ of dimension $d \geq 2$. Given $\kappa > 0$, there exists $\varepsilon = \varepsilon(E, \kappa) > 0$ such* *that the following holds for $\delta > 0$ sufficiently small. Let μ be a Borel probability measure on E. Assume that*

(i) μ is supported on $G_E(\delta^{-\varepsilon})$;
(ii) $\delta^{-\kappa} \leq \|\mu \boxplus P_\delta\|_2^2 \leq \delta^{-d+\kappa}$;
(iii) for every proper linear subspace $W < E$, $\forall \rho \geq \delta$, $\mu(W^{(\rho)}) \leq \delta^{-\varepsilon}\rho^\kappa$.

Then,

$$\|(\mu * \mu \boxminus \mu * \mu) \boxplus P_\delta\|_2 \leq \delta^\varepsilon \|\mu \boxplus P_\delta\|_2.$$

Remark 1. If $E = \mathbb{R}$ the same holds if condition (iii) is replaced by

$$\forall \rho \geq \delta, \forall x \in E,\ \mu(B_E(x,\rho)) \leq \delta^{-\varepsilon}\rho^\kappa. \tag{2.1}$$

Note also that when $\dim(E) \geq 2$, property (2.1) is implied by condition (iii).

Remark 2. We shall apply this proposition to measures that are not probability measures. It is clear that by making ε slightly smaller, the same statement holds for measures μ with total mass $\mu(E) \geq \frac{1}{2}$ or just $\mu(E) \geq \delta^\varepsilon$.

Proof. In this proof, the implied constants in the Landau or Vinogradov notation depend on the algebra structure of E as well as the choice of norm on it. We use the following rough comparison notation : for positive quantities f and g, we write $f \lessapprox g$ if $f \leq \delta^{-O(\varepsilon)} g$ and $f \sim g$ for $f \lessapprox g$ and $g \lessapprox f$. For instance, if $a \in G_E(\delta^{-\varepsilon})$ then $|\det_E(a)| \sim 1$.

To simplify notation, we shall also use the shorthand $\mu_\delta = \mu \boxplus P_\delta$. Now assume for a contradiction that the conclusion of the proposition does not hold, namely

$$\|(\mu * \mu \boxminus \mu * \mu) \boxplus P_\delta\|_2 \geq \delta^\varepsilon \|\mu_\delta\|_2. \tag{2.2}$$

Step 0: Compare the L^2-norms of $(\mu * \mu \boxminus \mu * \mu) \boxplus P_\delta$ and $\mu * \mu_\delta \boxminus \mu_\delta * \mu$.
For $x \in E$, write

$$\begin{aligned}
&(\mu * \mu \boxminus \mu * \mu \boxplus P_\delta)(x)\\
=&|B(0,\delta)|^{-1}\mu^{\otimes 4}\{(a,b,c,d) \mid ab - cd \in B_E(x,\delta)\}\\
\leq&|B(0,\delta)|^{-1}(\mu^{\otimes 4} \otimes P_\delta^{\otimes 2})\{(a,b,c,d,y,z) \mid a(b+y) - (c+z)d \in B_E(x,\delta^{1-2\varepsilon})\}\\
\lessapprox&|B(0,\delta^{1-2\varepsilon})|^{-1}(\mu^{\otimes 4} \otimes P_\delta^{\otimes 2})\{(a,b,c,d,y,z) \mid a(b+y) - (c+z)d \in B_E(x,\delta^{1-2\varepsilon})\}\\
=&(\mu * \mu_\delta \boxminus \mu_\delta * \mu \boxplus P_{\delta^{1-2\varepsilon}})(x).
\end{aligned}$$

Above at the sign $\leq$, we used the assumption that $\mathrm{Supp}(\mu) \subset B_E(0,\delta^{-\varepsilon})$. Therefore, by Young's inequality,

$$\|\mu * \mu \boxminus \mu * \mu \boxplus P_\delta\|_2 \lessapprox \|\mu * \mu_\delta \boxminus \mu_\delta * \mu \boxplus P_{\delta^{1-2\varepsilon}}\|_2 \leq \|\mu * \mu_\delta \boxminus \mu_\delta * \mu\|_2.$$

To conclude step 0, we deduce from the above and (2.2) that

$$\|\mu * \mu_\delta \boxminus \mu_\delta * \mu\|_2 \gtrapprox \|\mu_\delta\|_2. \tag{2.3}$$

Step 1: Discretize the measure μ using dyadic level sets.
For a subset $A \subset E$, we denote by $\mathcal{N}(A,\delta)$ the least number of balls of radius δ in E that is needed to cover A. By a δ-discretized set we mean a union of balls of radius δ. Note that if A is a δ-discretized set then $\mathcal{N}(A,\delta) \asymp_E \delta^{-d}|A|$. Is is easy to check that there exist δ-discretized sets $A_i \subset B_E(\delta^{-\varepsilon})$, $i \geq 0$ such that A_i is empty for $i \gg \log\frac{1}{\delta}$, and

$$\mu_\delta \ll \sum_{i\geq 0} 2^i\, \mathbb{1}_{A_i} \ll \mu_{3\delta} + 1. \tag{2.4}$$

Step 2: Pick a popular level in order to transform (2.3) into a lower bound on the additive energy between two δ-discretized sets.
We have

$$\mu * \mu_\delta \boxminus \mu_\delta * \mu = \iint_{E\times E} (\mathcal{D}_a * \mu_\delta) \boxminus (\mu_\delta * \mathcal{D}_b)\, \mathrm{d}\mu(a)\, \mathrm{d}\mu(b).$$

From the left inequality in (2.4),

$$\mu * \mu_\delta \boxminus \mu_\delta * \mu \ll \sum_{i,j\geq 0} 2^{i+j} \iint (\mathcal{D}_a * \mathbb{1}_{A_i}) \boxminus (\mathbb{1}_{A_j} * \mathcal{D}_b)\, \mathrm{d}\mu(a)\, \mathrm{d}\mu(b).$$

Observe that $\mathcal{D}_a * \mathbb{1}_{A_i} = |\det_E(a)|^{-1}\, \mathbb{1}_{aA_i}$ and $\mathbb{1}_{A_j} * \mathcal{D}_b = |\det_E(b)|^{-1}\, \mathbb{1}_{A_j b}$. Hence

$$\mu * \mu_\delta \boxminus \mu_\delta * \mu \ll \sum_{i,j\geq 0} 2^{i+j} \iint \frac{\mathbb{1}_{aA_i} \boxminus \mathbb{1}_{A_j b}}{|\det_E(a)\det_E(b)|}\, \mathrm{d}\mu(a)\, \mathrm{d}\mu(b).$$

By (2.3), the triangular inequality and the assumption that μ is supported on $G_E(\delta^{-\varepsilon})$,

$$\sum_{i,j\geq 0} 2^{i+j} \iint \|\mathbb{1}_{aA_i} \boxminus \mathbb{1}_{A_j b}\|_2\, \mathrm{d}\mu(a)\, \mathrm{d}\mu(b) \gtrsim \|\mu_\delta\|_2,$$

There are at most $O(\log\frac{1}{\delta})^2 \lesssim 1$ terms in this sum. Hence by the pigeonhole principle, there exist $i \geq 0$ and $j \geq 0$ such that

$$2^{i+j} \iint \|\mathbb{1}_{aA_i} \boxminus \mathbb{1}_{A_j b}\|_2\, \mathrm{d}\mu(a)\, \mathrm{d}\mu(b) \gtrsim \|\mu_\delta\|_2. \tag{2.5}$$

From now on we fix such i and j. By the right-hand inequality in (2.4), we find

$$2^i|A_i|^{1/2} = \|2^i\, \mathbb{1}_{A_i}\|_2 \ll \|\mu_{3\delta}\|_2 + 1 \ll \|\mu_\delta\|_2,$$

so that for all $a, b \in G_E(\delta^{-O(\varepsilon)})$,

$$\|2^i\, \mathbb{1}_{aA_i}\|_2 \lesssim \|\mu_\delta\|_2 \text{ and } \|2^j\, \mathbb{1}_{A_j b}\|_1 \lesssim 1. \tag{2.6}$$

Hence by Young's inequality,

$$\forall a, b \in G_E(\delta^{-O(\varepsilon)}),\ 2^{i+j}\|\mathbb{1}_{aA_i} \boxminus \mathbb{1}_{A_j b}\|_2 \lesssim \|\mu_\delta\|_2.$$

This combined with (2.5) implies that the set

$$B_0 = \{(a,b) \in G_E(\delta^{-O(\varepsilon)})^{\times 2} \mid 2^{i+j}\|\mathbb{1}_{aA_i} \boxminus \mathbb{1}_{A_j b}\|_2 \geq \delta^{O(\varepsilon)}\|\mu_\delta\|_2\}$$

has measure $\mu \otimes \mu(B_0) \gtrsim 1$. For $c = (a,b) \in B_0$, using (2.6), we find

$$\|\mathbb{1}_{aA_i} \boxminus \mathbb{1}_{A_j b}\|_2 \gtrsim |aA_i|^{1/2}|A_j b|,$$

and switching the role of aA_i and $A_j b$,

$$\|\mathbb{1}_{aA_i} \boxminus \mathbb{1}_{A_j b}\|_2 \gtrsim |aA_i||A_j b|^{1/2}.$$

Hence,

$$\|\mathbb{1}_{aA_i} \boxminus \mathbb{1}_{A_j b}\|_2^2 \gtrsim |aA_i|^{3/2}|A_j b|^{3/2}.$$

Note that aA_i and $A_j b$ are $\delta^{1-O(\varepsilon)}$-discretized sets. Hence the last inequality translates to

$$\mathcal{E}_\delta(aA_i, -A_j b) \gtrsim \mathcal{N}(aA_i, \delta)^{3/2}\mathcal{N}(A_j b, \delta)^{3/2} \tag{2.7}$$

where $\mathcal{E}_\delta$ denotes the additive energy at scale δ, as defined in [39, Section 6] or [16, Appendix A.1].
Step 3: Apply an argument of Bourgain [9, Proof of Theorem C] sometimes known as the additive-multiplicative Balog-Szemerédi-Gowers theorem.
We are going to use Rusza calculus. For subsets $A, A' \subset E$ we write $A \approx A'$ if

$$\mathcal{N}(A - A', \delta) \lesssim \mathcal{N}(A, \delta)^{1/2}\mathcal{N}(A', \delta)^{1/2}.$$

Ruzsa's triangular inequality and the Plünnecke-Ruzsa inequality [40, Chapters 2 & 6] can be summarized as : the relation $\approx$ is transitive[1], i.e. $A \approx A'$ and $A' \approx A''$ implies $A' \approx A''$.

By Tao's non-commutative version of the Balog-Szemerédi-Gowers lemma [39, Theorem 6.10] applied to (2.7), for every $c \in B_0$, there exists $A_c \subset A_i$ and $A'_c \subset A_j$ such that $\mathcal{N}(A_c, \delta) \gtrsim \mathcal{N}(A_i, \delta)$ and $\mathcal{N}(A'_c, \delta) \gtrsim \mathcal{N}(A_j, \delta)$ and

$$aA_c \approx A'_c b. \tag{2.8}$$

By taking δ-neighborhoods if necessary, we may assume that A_c and A'_c are δ-discretized sets. Write $X = A_i \times A_j \subset \mathbb{R}^{2d}$ and $X_c = A_c \times A'_c \subset X$. From the Cauchy-Schwarz inequality applied to the function $x \mapsto \int_{B_0} \mathbb{1}_{X_c}(x) \,\mathrm{d}\mu^{\otimes 2}(c)$, we infer that

$$\iint_{B_0 \times B_0} |X_c \cap X_d| \,\mathrm{d}\mu^{\otimes 2}(c) \,\mathrm{d}\mu^{\otimes 2}(d) \gtrsim |X|.$$

By the pigeonhole principle, there exists $c_\star \in B_0$ and $B_1 \subset B_0$ such that

$$\mu^{\otimes 2}(B_1) \gtrsim \mu^{\otimes 2}(B_0) \gtrsim 1$$

and for all $c \in B_1$, $|X_{c_\star} \cap X_c| \gtrsim |X|$. Abbreviate $A_{c_\star}$ as $A_\star$ and $A'_{c_\star}$ as $A'_\star$. We then have, for every $c \in B_1$,

$$\mathcal{N}(A_\star \cap A_c, \delta) \gtrsim \mathcal{N}(A_i, \delta) \text{ and } \mathcal{N}(A'_\star \cap A'_c, \delta) \gtrsim \mathcal{N}(A_j, \delta). \tag{2.9}$$

For $c = (a, b) \in B_1$, by the Rusza calculus and (2.8),

$$aA_c \approx A'_c b \approx aA_c.$$

Since $a \in G_E(\delta^{-\varepsilon})$, this implies

$$A_c \approx A_c.$$

Using (2.9) and the definition of the symbol $\approx$, we get $A_\star \cap A_c \approx A_c$ and for the same reason $A_\star \cap A_c \approx A_\star$. Hence

$$aA_\star \approx a(A_\star \cap A_c) \approx aA_c \approx A'_c b \approx (A'_\star \cap A'_c) b \approx A'_\star b.$$

On the other hand, writing $c_\star = (a_\star, b_\star)$, we have $a_\star A_\star \approx A'_\star b_\star$. Hence $a_\star A_\star b_\star^{-1} b \approx A'_\star b$ and then $a_\star A_\star b_\star^{-1} b \approx aA_\star$ and finally

$$\mathcal{N}(a_\star^{-1} a A_\star - A_\star b_\star^{-1} b, \delta) \lesssim \mathcal{N}(A_\star, \delta), \quad \forall (a,b) \in B_1 \cup \{(a_\star, b_\star)\}. \tag{2.10}$$

<u>Step 5:</u> Apply the sum-product theorem stated below as Proposition 2.3.
We claim that the assumptions of Proposition 2.3 are satisfied by the set $A_\star$, the set B_1 and the measure μ for the parameters $\kappa/2$ in the place of κ and $O(\varepsilon)$ in the place of ε. Indeed, using Young's inequality and remembering (2.6), we obtain

$$\|\mu_\delta\|_2 \lesssim 2^{i+j} \|\mathbb{1}_{a_\star A_i} \boxminus \mathbb{1}_{A_j b_\star}\| \leq 2^i |a_\star A_i|^{1/2} 2^j |A_j b_\star| \lesssim \|\mu_\delta\|_2.$$

Hence $2^i |A_i|^{1/2} \sim \|\mu_\delta\|_2$ and $2^j |A_j| \sim 1$. Inversing the roles of A_i and A_j, we get also $2^i |A_i| \sim 1$. Thus,

$$|A_i| \sim \|\mu_\delta\|_2^{-2} \text{ and } 2^i \sim \|\mu_\delta\|_2^2. \tag{2.11}$$

Hence

$$|A_\star| \lesssim |A_i| \lesssim \|\mu_\delta\|_2^{-2} \leq \delta^\kappa,$$

which implies, as $A_\star$ is δ-discretized,

$$\mathcal{N}(A_\star, \delta) \lesssim \delta^{-d+\kappa}.$$

Moreover, let $\rho \geq \delta$ and $x \in E$ and let $B = B(x, \rho)$. Since $\mu_{3\delta} = \mu \boxplus P_{3\delta}$, inequality (2.1) implies

$$\mu_{3\delta}(B) \lesssim \rho^\kappa.$$

[1] Strictly speaking, $\approx$ is not relation, because it involves an implicit constant in the $\lesssim$ notation.

By (2.11) and (2.4),

$$\frac{|A_i \cap B|}{|A_i|} \lesssim 2^i |A_i \cap B| \ll \mu_{3\delta}(B)\big).$$

But $A_\star \subset A_i$ and $|A_\star| \gtrsim |A_i|$, hence

$$\frac{|A_\star \cap B|}{|A_\star|} \lesssim \frac{|A_i \cap B|}{|A_i|} \lesssim \rho^\kappa.$$

It follows that for all $\rho \geq \delta$,

$$\mathcal{N}(A_\star, \rho) \gtrsim \rho^{-\kappa}.$$

The verification of the other assumptions in Proposition 2.3 are straightforward, so we can apply Proposition 2.3, which leads to a contradiction to (2.10) when $\varepsilon > 0$ is chosen small enough. $\square$

2.2. **A sum-product theorem.** In the proof of L^2-flattening, we used the following result.

Proposition 2.3 (Sum-product estimate in simple algebras)**.** *Let E be a normed finite-dimensional simple algebra over $\mathbb{R}$ of dimension $d \geq 2$. Given $\kappa > 0$, there exists $\varepsilon = \varepsilon(E, \kappa) > 0$ such that the following holds for every $\delta > 0$ sufficiently small. Let A be a subset of E, μ a probability measure on E, and B a subset of $E \times E$. Assume*

(i) $A \subset B_E(0, \delta^{-\varepsilon})$;
(ii) $\forall \rho \geq \delta$, $\mathcal{N}(A, \rho) \geq \delta^\varepsilon \rho^{-\kappa}$;
(iii) $\mathcal{N}(A, \delta) \leq \delta^{-(d-\kappa)}$;
(iv) *μ is supported on* $G_E(\delta^{-\varepsilon})$;
(v) *for every proper linear subspace* $W < E$, $\forall \rho \geq \delta$, $\mu(W^{(\rho)}) \leq \delta^{-\varepsilon}\rho^\kappa$;
(vi) $\mu \otimes \mu(B) \geq \delta^\varepsilon$.

Then for every $a_\star, b_\star \in G_E(\delta^{-\varepsilon})$, there exists $(a, b) \in B \cup \{(a_\star, b_\star)\}$ such that

$$\mathcal{N}(a_\star^{-1} a A + A b_\star^{-1} b, \delta) \geq \delta^{-\varepsilon} \mathcal{N}(A, \delta).$$

The idea of the proof is to consider the action of $E \times E$ on E by left and right multiplication and to apply a sum-product theorem [25, Theorem 3] for irreducible linear actions due to the first author. For the reader's convenience, let us recall the statement of the latter.

Theorem 2.4 (Sum-product theorem for irreducible linear actions)**.** *Given a positive integer d and a real number $\kappa > 0$ there exists $\varepsilon = \varepsilon(d, \kappa) > 0$ such that the following holds for $\delta > 0$ sufficiently small. Let X be a subset of the Euclidean space $\mathbb{R}^d$ and $\Phi \subset \mathrm{End}(\mathbb{R}^d)$ a subset of linear endomorphisms. Assume*

(i) $X \subset B_{\mathbb{R}^d}(0, \delta^{-\varepsilon})$;
(ii) *for all* $\rho \geq \delta$, $\mathcal{N}(X, \rho) \geq \delta^\varepsilon \rho^{-\kappa}$;
(iii) $\mathcal{N}(X, \delta) \leq \delta^{-(d-\kappa)}$;
(iv) $\Phi \subset B_{\mathrm{End}(\mathbb{R}^d)}(0, \delta^{-\varepsilon})$;
(v) *for all* $\rho \geq \delta$, $\mathcal{N}(\Phi, \rho) \geq \delta^\varepsilon \rho^{-\kappa}$;
(vi) *for every proper linear subspace $W \subset \mathbb{R}^d$, there is $\varphi \in \Phi$ and $w \in B_W(0,1)$ such that* $d(\varphi w, W) \geq \delta^\varepsilon$.

Then

$$\mathcal{N}(X + X, \delta) + \max_{\varphi \in \Phi} \mathcal{N}(X + \varphi X, \delta) \geq \delta^{-\varepsilon} \mathcal{N}(X, \delta).$$

Here, of course, φX denote the set $\{\varphi x \mid x \in X\}$

Proof of Proposition 2.3. In this proof the implied constants in the Vinogradov or Landau notation may depend on E. We may assume without loss of generality that $B \subset \mathrm{Supp}(\mu) \times \mathrm{Supp}(\mu)$. This implies that for all $(a,b) \in B$,

$$\|a\|, \|a^{-1}\|, \|b\|, \|b^{-1}\| \leq \delta^{-\varepsilon},$$

and consequently the multiplication on the left or right by a or b is a $\delta^{-O(\varepsilon)}$-bi-Lipschitz endomorphism of E.

For (a,b) in B, define $\varphi(a,b) \in \mathrm{End}(E)$ by

$$\forall x \in E,\ \varphi(a,b)x = -a^{-1}a_\star x b_\star^{-1} b.$$

We would like to apply the previous theorem to $X = A$ and $\Phi = \{\varphi(a,b) \in \mathrm{End}(E) \mid a, b \in B\}$. We claim that the assumptions of Theorem 2.4 hold with $O(\varepsilon)$ in the place of ε. Hence there is $\varepsilon_1 > 0$ such that when $\varepsilon > 0$ is small enough, we have either

$$\mathcal{N}(A + A, \delta) \geq \delta^{-\varepsilon_1}\mathcal{N}(A, \delta)$$

in which case we are done, or there exists $(a,b) \in B$ such that

$$\mathcal{N}(A + \varphi(a,b)A, \delta) \geq \delta^{-\varepsilon_1}\mathcal{N}(A, \delta).$$

In the latter case we conclude by multiplying the set above by $a_\star^{-1}a$ on the left,

$$\mathcal{N}(a_\star^{-1}aA - Ab_\star^{-1}b, \delta) \geq \delta^{O(\varepsilon)}\mathcal{N}(A + \varphi(a,b)A, \delta) \geq \delta^{-\varepsilon_1 + O(\varepsilon)}\mathcal{N}(A, \delta).$$

It remains to check the assumptions in Theorem 2.4. Items (i)–(iv) are immediate. To check the remaining assumptions, write, for $b \in E$,

$$B_1(b) = \{a \in E \mid (a,b) \in B\}.$$

By assumption (vi) of the proposition we are trying to prove, we can pick $b_0 \in E$ such that

$$\mu(B_1(b_0)) \geq \delta^{\varepsilon}.$$

From the inequalities

$$\|a - a'\| \ll \|a\|\|a'\|\|a'^{-1} - a^{-1}\|$$

and

$$\|a'^{-1} - a^{-1}\| = \|\big(\varphi(a,b_0) - \varphi(a',b_0)\big)(a_\star^{-1}b_0^{-1}b_\star)\| \leq \|\varphi(a,b_0) - \varphi(a',b_0)\|\|a_\star^{-1}b_0^{-1}b_\star\|,$$

we see that the map $a \mapsto \varphi(a,b_0)$ is $\delta^{-O(\varepsilon)}$-bi-Lipschitz on $B_1(b_0)$. Thus item (v) follows from assumption (v) of Proposition 2.3.

Finally, assume for contradiction that item (vi) fails with $\delta^{C\varepsilon}$ in the place of δ^{ε}. Namely, there is a linear subspace $W_0 \subset \mathbb{R}^d$ of intermediate dimension $0 < k < d$ such that $\forall (a,b) \in B$, $d(W_0, \varphi(a,b)W_0) \leq \delta^{C\varepsilon}$, where d denotes the distance on the the Grassmannian $\mathrm{Grass}(k,d)$ of k-planes in $\mathbb{R}^d$ defined by

$$d(W, W') = \min_{w \in B_W(0,1)} d(w, W') = \min_{w' \in B_{W'}(0,1)} d(w', W).$$

In particular, for $a, a' \in B_1(b_0)$, we have

$$d(W_0, \varphi(a,b_0)W_0) \leq \delta^{C\varepsilon} \text{ and } d(W_0, \varphi(a',b_0)W_0) \leq \delta^{C\varepsilon}.$$

Multiplying the second inequality on the left by $a^{-1}a'$, we obtain

$$d(a^{-1}a'W_0, \varphi(a,b)W_0) \leq \delta^{(C - O(1))\varepsilon}.$$

By the triangular inequality,

$$d(W_0, a^{-1}a'W_0) \leq \delta^{(C-O(1))\varepsilon},$$

which means

(2.12)
$$\forall g \in a^{-1}B_1(b_0),\ \forall w \in W_0,\ d(gw, W_0) \leq \|gw\| d(gW_0, W_0) \leq \delta^{(C-O(1))\varepsilon}\|w\|.$$

Observe that the assumption (v) of Proposition 2.3 implies that the subset $B_1(b_0) \subset E$ is $\delta^{O(\varepsilon)}$-away from linear subspaces. Hence so is the subset $a^{-1}B_1(b_0)$. Using [25, Lemma 8], we obtain from (2.12),

$$\forall x \in B_E(0,1),\ \forall w \in B_{W_0}(0,1),\ d(xw, W_0) \leq \delta^{(C-O(1))\varepsilon}.$$

We can do the same argument for the right multiplication. Thus, similarly,

$$\forall x \in B_E(0,1),\ \forall w \in B_{W_0}(0,1),\ d(wx, W_0) \leq \delta^{(C-O(1))\varepsilon}.$$

Consider the map $f\colon \operatorname{Grass}(k,d) \to \mathbb{R}$ defined by

$$f(W) = \iint_{B_E(0,1)\times B_W(0,1)} \big(d(xw,W) + d(wx,W)\big)\,\mathrm{d}x\,\mathrm{d}w.$$

On the one hand, from the above, $f(W_0) \leq \delta^{(C-O(1))\varepsilon}$. On the other hand, f is continuous and defined on a compact set. It never vanishes for the reason that a zero of f must be a two-sided ideal of E contradicting the simplicity of E. Hence f has a positive minimum on $\operatorname{Grass}(k,d)$. We obtain a contradiction if C is chosen large enough, proving our claim regarding item (vi). □

2.3. **Fourier decay for multiplicative convolutions.** The goal here is to prove Theorem 2.1 using iteratively the L^2-flattening lemma proved above.

Let E be any finite-dimensional real algebra. The Fourier transform of a finite Borel measure μ on E is the function on the dual E^* given by

$$\forall \xi \in E^*,\ \hat{\mu}(\xi) = \int_E e(\xi x)\,\mathrm{d}\mu(x).$$

where $e(t) = e^{2\pi i t}$ for $t \in \mathbb{R}$, and we simply write $E^* \times E \to \mathbb{R};\ (\xi, x) \mapsto \xi x$ for the duality pairing. The product on E yields a natural right action of E on E^* given by

$$\forall \xi \in E^*,\ x \in E,\ y \in E, \quad (\xi y)(x) = \xi(yx),$$

and for finite Borel measures μ and ν on E, the Fourier transform of their multiplicative convolution is given by

$$\widehat{\mu * \nu}(\xi) = \int \hat{\nu}(\xi y)\,\mathrm{d}\mu(y). \tag{2.13}$$

The idea of the proof of Theorem 2.1 is to iterate Proposition 2.2 to get a measure with small L^2-norm, and then to get the desired Fourier decay by convolving one more time. Two technical issues arise. First, after each iteration, the measure we obtain does not necessarily satisfy the non-concentration property required by Proposition 2.2. To settle this, at each step, we truncate the measure to restrict the support on well-invertible elements. Second, the measure we obtain in the end of the iteration is not an additive convolution of a multiplicative convolution of μ but some measure obtained from μ through successive multiplicative and additive convolutions. To conclude we need to clarify relation between the Fourier transforms of these measures. This is settled in Lemma 2.7.

Lemma 2.5. *Let E be a finite-dimensional normed algebra over $\mathbb{R}$ and μ a Borel probability measure on E such that for some $\tau, \varepsilon > 0$,*

(i) $\mu\big(E \setminus B_E(0, \delta^{-\varepsilon})\big) \leq \delta^{\tau}$*;*
(ii) for every $x \in E$, $\mu(x + S_E(\delta^{\varepsilon})) \leq \delta^{\tau}$*;*
(iii) for every $\rho \geq \delta$ *and every proper affine subspace* $W \subset E$, $\mu(W^{(\rho)}) \leq \delta^{-\varepsilon}\rho^{\kappa}$.

Set $\mu_1 = \mu_{|B_E(0,\delta^{-\varepsilon})}$ *and define recursively for integer* $k \geq 1$,

$$\eta_k = \mu_{k|E\setminus S_E(\delta^{2^k\varepsilon})}$$

and

$$\mu_{k+1} = \eta_k * \eta_k \boxminus \eta_k * \eta_k.$$

Then we have for $k \geq 1$,

$$\mu_k(E) \geq 1 - O_k(\delta^\tau) \tag{2.14}$$

$$\mathrm{Supp}(\mu_k) \subset B_E(0, \delta^{-O_k(\varepsilon)}) \tag{2.15}$$

$$\forall x \in E,\ \mu_k(x + S_E(\delta^{2^k\varepsilon})) \leq \delta^\tau \tag{2.16}$$

$$\forall \rho \geq \delta, \forall W \subset E \textit{ proper affine subspace},\ \mu_k(W^{(\rho)}) \leq \delta^{-O_k(\varepsilon)}\rho^\kappa. \tag{2.17}$$

As a consequence, the same holds for η_k in the place of μ_k.

Proof. The proof goes by induction on k.
The result is clear for $k = 1$, by assumption on μ. Assume (2.14)–(2.17) true for some $k \geq 1$, so that the same holds for η_k. Then (2.14) and (2.15) for $k+1$ follow immediately.

Let us prove (2.16) for $k+1$. Let $x \in E$. Since $\mu_{k+1} = \eta_k * \eta_k \boxminus \eta_k * \eta_k$,

$$\mu_{k+1}\big(x+S_E(\delta^{2^{k+1}\varepsilon})\big) = \iint \eta_k\{y \in E \mid |\det_E(yz-w-x)| \leq \delta^{2^{k+1}\varepsilon}\}\, \mathrm{d}\eta_k(z)\, \mathrm{d}(\eta_k*\eta_k)(w)$$

Note that for $z \in \mathrm{Supp}(\eta_k)$, by definition, $|\det_E(z)| \geq \delta^{2^k\varepsilon}$. Hence $|\det_E(yz - w - x)| \leq \delta^{2^{k+1}\varepsilon}$ implies $y - (w+x)z^{-1} \in S_E(\delta^{2^k\varepsilon})$. Therefore, by induction hypothesis (2.16)

$$\mu_{k+1}\big(x + S_E(\delta^{2^{k+1}\varepsilon})\big) \leq \max_{z\in\mathrm{Supp}(\eta_k),\, w\in E} \eta_k\big((w+x)z^{-1} + S_E(\delta^{2^k\varepsilon})\big) \leq \delta^\tau.$$

Finally, let us prove (2.17) for $k+1$. Let $\rho \geq \delta$ and let W be a proper affine subspace of E. We have as above

$$\mu_{k+1}\big(W^{(\rho)}\big) \leq \max_{z\in\mathrm{Supp}(\eta_k),\, w\in E} \eta_k\big((w + W^{(\rho)})z^{-1}\big).$$

For all $z \in \mathrm{Supp}(\eta_k)$, we have $|\det_E(z)| \geq \delta^{O_k(\varepsilon)}$ and by the induction hypothesis $\|z\| \leq \delta^{-O_k(\varepsilon)}$. Hence $\|z^{-1}\| \leq \delta^{-O_k(\varepsilon)}$. Thus

$$(w + W^{(\rho)})z^{-1} = wz^{-1} + Wz^{-1} + B_E(0,\rho)z^{-1} \subset wz^{-1} + Wz^{-1} + B_E(0, \delta^{-O_k(\varepsilon)}\rho),$$

which is nothing but the $(\delta^{-O_k(\varepsilon)}\rho)$-neighborhood of another proper affine subspace. Hence by induction hypothesis (2.17),

$$\mu_{k+1}\big(W^{(\rho)}\big) \leq \delta^{-O_k(\varepsilon)}\rho^\kappa.$$

This finishes the proof of the induction step and that of the lemma. □

Lemma 2.6. *Let E be normed algebra over $\mathbb{R}$ of dimension d. Let μ and ν be Borel probability measures on E. Assume*

(i) $\|\mu \boxplus P_\delta\|_2^2 \leq \delta^\kappa$,

(ii) for every $\rho \geq \delta$ and every proper affine subspace $W \subset E$, $\nu(W^{(\rho)}) \leq \delta^{-\varepsilon}\rho^\kappa$.

Then for $\xi \in E^*$ with $\delta^{-1+\varepsilon} \leq \|\xi\| \leq \delta^{-1-\varepsilon}$,

$$|\widehat{\mu * \nu}(\xi)| \leq \delta^{\frac{2\kappa}{d+3} - O(\varepsilon)}$$

Proof. This is a slightly more general form of [10, Theorem 7]. The proof is essentially the same. A detailed proof is implicitly contained in [27, Lemma 2.11]. □

For a finite Borel measure ν on E and an integer $\ell \geq 1$, we denote by

$$\nu^{\boxplus \ell} = \underbrace{\nu \boxplus \cdots \boxplus \nu}_{\ell \text{ times}}$$

the ℓ-fold additive convolution of ν.

Lemma 2.7. *Let E be a finite-dimensional real algebra. Let $\ell \geq 1$ be an integer, ν a Borel probability measure on E, and set*

$$\mu = \nu^{\boxplus \ell} \boxminus \nu^{\boxplus \ell}.$$

Then for every integer $m \geq 1$, for every $\xi \in E^*$, the Fourier coefficient $\widehat{\mu^{*m}}(\xi)$ is real and

$$|\widehat{\nu^{*m}}(\xi)|^{(2\ell)^m} \leq \widehat{\mu^{*m}}(\xi). \tag{2.18}$$

The same inequality also holds for a finite Borel measure with total mass $\nu(E) \leq 1$.

Proof. We proceed by induction on m. For $m = 1$,

$$\hat{\mu}(\xi) = |\hat{\nu}(\xi)|^{2\ell}.$$

Assume then that (2.18) is true for some $m \geq 1$. By (2.13), the Hölder inequality and the induction hypothesis,

$$\begin{aligned} |\widehat{\nu^{*(m+1)}}(\xi)|^{(2\ell)^m} &= \Big| \int \widehat{\nu^{*m}}(\xi y) \, \mathrm{d}\nu(y) \Big|^{(2\ell)^m} \\ &\leq \int |\widehat{\nu^{*m}}(\xi y)|^{(2\ell)^m} \, \mathrm{d}\nu(y) \\ &\leq \int \widehat{\mu^{*m}}(\xi y) \, \mathrm{d}\nu(y) \\ &= \iint e(\xi y x) \, \mathrm{d}\mu^{*m}(x) \, \mathrm{d}\nu(y) \end{aligned}$$

Taking the 2ℓ-th power and using again the Hölder inequality and (2.13), we obtain

$$\begin{aligned} &|\widehat{\nu^{*(m+1)}}(\xi)|^{(2\ell)^{m+1}} \\ \leq& \Big| \iint e(\xi y x) \, \mathrm{d}\nu(y) \, \mathrm{d}\mu^{*m}(x) \Big|^{2\ell} \\ \leq& \int \Big| \int e(\xi y x)) \, \mathrm{d}\nu(y) \Big|^{2\ell} \mathrm{d}\mu^{*m}(x) \\ =& \iiint e\big(\xi(y_1 + \cdots + y_\ell - y_{\ell+1} - \cdots - y_{2\ell})x\big) \, \mathrm{d}\nu(y_1) \ldots \mathrm{d}\nu(y_{2\ell}) \, \mathrm{d}\mu^{*m}(x) \\ =& \iint e(\xi z x) \, \mathrm{d}\mu(z) \, \mathrm{d}\mu^{*m}(x) \\ =& \widehat{\mu^{*(m+1)}}(\xi) \end{aligned}$$

This proves the induction step and finishes the proof of (2.18). If ν is a finite Borel measure with $\nu(E) \leq 1$, we may apply (2.18) to the probability measure $\nu(E)^{-1}\nu$, which yields

$$\Big(\frac{|\widehat{\nu^{*m}}(\xi)|}{\nu(E)^m} \Big)^{(2\ell)^m} \leq \frac{\widehat{\mu^{*m}}(\xi)}{\nu(E)^{2\ell m}}.$$

□

Proof of Theorem 2.1. Let $d = \dim(E)$. Let $\varepsilon_1 > 0$ be the constant given by Proposition 2.2 applied to the parameter $\kappa/2$. Define $s_{\max} = \lceil \frac{d}{\varepsilon_1} \rceil$. First, remark that by the non-concentration assumption, $\|\mu \boxplus P_\delta\|_\infty \leq \delta^{-d+\kappa-\varepsilon}$. Hence

$$\|\mu \boxplus P_\delta\|_2^2 \leq \|\mu \boxplus P_\delta\|_1 \|\mu \boxplus P_\delta\|_\infty \leq \delta^{-d+\kappa/2}$$

if we choose $\varepsilon \leq \kappa/2$. For $k = 1, \ldots, s_{\max}$, let μ_k and η_k be defined as in Lemma 2.5. Since $k \leq s_{\max}$ is bounded, the implied constants in the $O_k(\varepsilon)$ notations in Lemma 2.5 can be chosen uniformly over k. Thus, when $\varepsilon > 0$ is sufficiently small, Lemma 2.5 allows us to apply Proposition 2.2 and Remark 2 after it to the measure η_k for each $k = 1, \dots, s_{\max}$. Thus, either $\|\eta_k \boxplus P_\delta\|_2^2 \leq \delta^{\kappa/2}$ or

$$\|\eta_{k+1} \boxplus P_\delta\|_2^2 \leq \|\mu_{k+1} \boxplus P_\delta\|_2^2 \leq \delta^{\varepsilon_1}\|\eta_k \boxplus P_\delta\|_2^2.$$

We deduce that there exists $s \in \{1, \dots, s_{\max}\}$ such that

$$\|\eta_s \boxplus P_\delta\|_2^2 \leq \delta^{\kappa/2}.$$

Remembering (2.17), we apply Lemma 2.6 to obtain, for all ξ with $\|\xi\| = \delta^{-1}$,

$$|\widehat{\eta_s * \eta_s}(\xi)| \leq \delta^{\kappa/(2d+6)} \leq \delta^\tau.$$

Here we assumed ε sufficiently small compared to κ/d.

For $k = 1, \dots, s$, apply Lemma 2.7 with $\ell = 1$ and $m = 2^k$ to $\mu_{s-k+1} = \eta_{s-k}^{*2} \boxminus \eta_{s-k}^{*2}$. We obtain

$$\left|\widehat{\eta_{s-k}^{*2^{k+1}}}(\xi)\right|^{2^{2^k}} \leq \left|\widehat{\mu_{s-k+1}^{*2^k}}(\xi)\right|.$$

Moreover, by (2.16), μ_{s-k+1} differs from η_{s-k+1} by a measure of total mass at most δ^τ. Hence

$$\left|\widehat{\mu_{s-k+1}^{*2^k}}(\xi)\right| \leq \left|\widehat{\eta_{s-k+1}^{*2^k}}(\xi)\right| + O_k(\delta^\tau).$$

From the above, we deduce using a simple recurrence that for all $k = 1, \dots, s$,

$$\left|\widehat{\mu_{s-k+1}^{*2^k}}(\xi)\right| \ll_k \delta^{\tau/O_k(1)}.$$

In particular,

$$\left|\widehat{\mu_1^{*2^s}}(\xi)\right| \ll_s \delta^{\varepsilon\tau},$$

which allows to conclude since μ differs from μ_1 by a measure of total mass at most δ^τ. □

3. Non-concentration in subvarieties

Our goal here is to prove that the law of a large random matrix product satisfies some regularity conditions.

Throughout this section, unless otherwise stated, μ denotes a probability measure on $\mathrm{SL}_d(\mathbb{Z})$. As in the introduction, Γ denotes the subsemigroup generated by $\mathrm{Supp}(\mu)$, $\mathbb{G} < \mathrm{SL}_d$ is the Zariski closure of Γ, and $G = \mathbb{G}(\mathbb{R})$ its group of $\mathbb{R}$-rational points. We also let E denote the subalgebra of $\mathcal{M}_d(\mathbb{R})$ generated by G, and fix a norm on the space of all polynomial functions on E. We shall prove two non-concentration statements for the distribution of a random matrix product. The first one shows that the law μ^{*n} at time n of the random matrix product is not concentrated near affine subspaces of the algebra E.

Proposition 3.1 (Non-concentration on affine subspaces). *Let μ be a probability measure on $\mathrm{SL}_d(\mathbb{Z})$, for some $d \geq 2$. Denote by Γ the subsemigroup generated by μ, and by $\mathbb{G}$ the Zariski closure of Γ in SL_d. Assume that:*

(a) The measure μ has a finite exponential moment;
(b) The action of Γ on $\mathbb{R}^d$ is irreducible;
(c) The algebraic group $\mathbb{G}$ is Zariski connected.

There exists $\kappa > 0$ such that for every proper affine subspace $W \subset E$,

$$\forall n \geq 1,\ \forall \rho \geq e^{-n}, \quad \mu^{*n}(\{g \in \Gamma \mid d(g, W) \leq \rho\|g\|\}) \ll_\mu \rho^\kappa.$$

The second result concerns general subvarieties of the algebra E, with the caveat that we have to replace μ^{*n} by an additive convolution power of itself, to avoid some obstructions. It is also worth noting that the quantification of the non-concentration is slightly weaker than in the case of affine subspaces.

Proposition 3.2 (Non-concentration on subvarieties)**.** *Let μ be a probability measure on $\mathrm{SL}_d(\mathbb{Z})$, for some $d \geq 2$. Let Γ denote the subsemigroup generated by μ, $\mathbb{G}$ the Zariski closure of Γ in SL_d, and λ_1 the top Lyapunov exponent of μ. Assume that:*

(a) The measure μ has a finite exponential moment;
(b) The action of Γ on $\mathbb{R}^d$ is irreducible;
(c) The algebraic group $\mathbb{G}$ is Zariski connected.

Given an integer $D \geq 1$ and given $\omega > 0$, there exists $c > 0$ and $n_0 \in \mathbb{N}$ such that the following holds. Let $f \colon E \to \mathbb{R}$ be a polynomial function of degree D. Writing f_D for its degree D homogeneous part, we have

$$\forall k \geq \dim(E),\ \forall n \geq n_0, \quad (\mu^{*n})^{\boxplus k}\big(\{x \in E \mid |f(x)| \leq e^{(D\lambda_1 - \omega)n}\|f_D\|\}\big) \leq e^{-cn}.$$

These two propositions will allow us to apply the results of the previous section to the law μ^{*n} at time n of an irreducible random walk on $\mathrm{SL}_d(\mathbb{Z})$. This will yield Theorem 3.19 below.

Non-concentration estimates for subvarieties can sometimes be obtained by some linearization techniques, as is done in [1]. But this approach does not seem to yield a uniform statement for subvarieties of bounded degree, which is crucial for our application.

The argument developed in this section relies on the spectral gap property modulo primes for finitely generated subgroups of $\mathbb{H}(\mathbb{Z})$, where $\mathbb{H}$ is a semisimple $\mathbb{Q}$-subgroup of SL_n, a still rather recent result obtained by Salehi Golsefidy and Varjú [38] after several important works in this direction, starting with Helfgott [29], followed by Bourgain and Gamburd [12], Breuillard, Green and Tao [18], and Pyber-Szabó [36].

3.1. **Prelude : Expansion in semisimple groups.** Since elements in Γ have integer coefficients, $\mathbb{G}$ is defined over $\mathbb{Q}$, so we may choose a set of defining polynomials with coefficients in $\mathbb{Z}$. Given a prime number p, this allows to consider the variety $\mathbb{G}_p$ defined over $\mathbb{F}_p$ by the reduction modulo p of the polynomials defining $\mathbb{G}$.

On the space $\ell^2(\mathbb{G}_p(\mathbb{F}_p))$ of square-integrable functions on $\mathbb{G}_p(\mathbb{F}_p)$, we shall consider the convolution operator

$$\begin{array}{rccc} T_\mu\colon & \ell^2(\mathbb{G}_p(\mathbb{F}_p)) & \to & \ell^2(\mathbb{G}_p(\mathbb{F}_p)) \\ & f & \mapsto & \mu * f \end{array}$$

Let $\ell^2_0(\mathbb{G}_p(\mathbb{F}_p)) \subset \ell^2(\mathbb{G}_p(\mathbb{F}_p))$ denote the subspace of functions on $\mathbb{G}_p(\mathbb{F}_p)$ having zero mean.

The theorem we shall need is the following; up to some minor modifications, it appears in Salehi Golsefidy and Varjú [38, Theorem 1].

Theorem 3.3 (Spectral gap theorem)**.** *Let $d \geq 2$ and let μ be a probability measure on $\mathrm{SL}_d(\mathbb{Z})$ such that the Zariski closed subgroup $\mathbb{G}$ generated by μ is semisimple. Then there exists a constant $c > 0$ and an integer k such that for every prime number p,*

$$\|T_\mu^k\|_{\ell^2_0(\mathbb{G}_p(\mathbb{F}_p))} \leq 1 - c.$$

Remark 3. Another way to state the above theorem is to say that the spectral radius of the operator T_μ restricted to $\ell^2_0(\mathbb{G}_p(\mathbb{F}_p))$ is bounded above by $1 - c$, for every prime number p.

For completeness, we now explain how to derive the above theorem from [38, Theorem 1]. The argument uses the following two lemmata.

Lemma 3.4. *Let Γ be a subsemigroup of $\mathrm{SL}_d(\mathbb{R})$ whose Zariski closure $\mathbb{G}$ is semisimple. There exists a finite subset $S \subset \Gamma$ such that the semigroup generated by S is Zariski dense in $\mathbb{G}$.*

Proof. For any finite subset $S \subset \Gamma$, denote by $Z_e(S)$ the identity component of the Zariski closure of the semigroup generated by S. Let $S_0 \subset \Gamma$ be a finite subset such that $\mathbb{H} := Z_e(S_0)$ has maximal dimension among these subgroups. Then $\mathbb{H}$ is also maximal for the order of inclusion, because $Z_e(S_0) \subset Z_e(S_0 \cup S)$ are both irreducible subvarieties and have the same dimension.

In particular, for any $\gamma \in \Gamma$, $\gamma\mathbb{H}\gamma^{-1} = Z_e(\gamma S_0 \gamma^{-1}) \subset \mathbb{H}$. Hence $\mathbb{H}$ is a normal subgroup in $\mathbb{G}$, since Γ is Zariski dense. By [7, Theorem 6.8 and Corollary 14.11], the quotient $\mathbb{G}/\mathbb{H}$ is a semisimple linear algebraic group such that the projection $\pi\colon \mathbb{G} \to \mathbb{G}/\mathbb{H}$ is a morphism of algebraic groups.

Moreover, for any $\gamma \in \Gamma$, there is $k \geq 1$ such that $\gamma^k \in Z_e(\{\gamma\}) \subset \mathbb{H}$. Hence the image $\pi(\Gamma)$ of Γ in $\mathbb{G}/\mathbb{H}$ is a torsion group. By the Jordan-Schur theorem [37, Theorem 8.31], $\pi(\Gamma)$ is virtually abelian. But $\pi(\Gamma)$ is Zariski dense in $\mathbb{G}/\mathbb{H}$. Hence the Zariski closure of any of its subgroups of finite index contains the identity component of $\mathbb{G}/\mathbb{H}$. Therefore, the identity component of $\mathbb{G}/\mathbb{H}$ is both semisimple and abelian, hence trivial. It follows that $\mathbb{G}/\mathbb{H}$ is finite. Thus, by adding a finite number of elements to S_0, we can make sure that S_0 generates a Zariski dense subsemigroup in $\mathbb{G}$. □

Lemma 3.5. *Let $\mathbb{G}$ be a connected semisimple algebraic group defined over $\mathbb{Q}$. Let $S \subset \mathbb{G}(\mathbb{Q})$ be a finite subset which generates a Zariski dense subgroup. Then there exists $k \geq 1$ such that the symmetric set $S^{-k}S^k$ also generates a Zariski dense subgroup in $\mathbb{G}$.*

Proof. By a result of Nori [35, Theorem 5.2], the group Γ generated by S is dense in some open subgroup Ω of $\mathbb{G}(\mathbb{Q}_p)$, for some prime number p. In other words, the union $\bigcup_{k\geq 1} S^k$ is dense in Ω. Since the Lie algebra $\mathfrak{g}$ of Ω satisfies $[\mathfrak{g},\mathfrak{g}] = \mathfrak{g}$, the derived subgroup $[\Omega,\Omega]$ contains an open subgroup Ω'; then, the increasing sequence of subsets $(S^kS^{-k})_{k\geq 1}$ gets arbitrarily dense in Ω'.

On the other hand, there exists a neighborhood U of the identity in $\mathbb{G}(\mathbb{Q}_p)$ and $\delta > 0$, such that if $\mathbb{H}$ is an algebraic subgroup of $\mathbb{G}$, then $\mathbb{H}(\mathbb{Q}_p)$ is not δ-dense in U. Indeed, the Lie algebra $\mathfrak{h}$ of $\mathbb{H}(\mathbb{Q}_p)$ satisfies $\dim_{\mathbb{Q}_p} \mathfrak{h} < \dim_{\mathbb{Q}_p} \mathfrak{g}$, and we can distinguish two cases. If the normalizer $N(\mathfrak{h})$ of $\mathfrak{h}$ in $\mathbb{G}(\mathbb{Q}_p)$ does not contain an open subgroup, then we conclude using [20, Lemma 2.2], which is still valid over $\mathbb{Q}_p$. Otherwise, $\mathfrak{h}$ is an ideal in $\mathfrak{g}$, so that $\mathbb{H}$ is a sum of simple factors of $\mathbb{G}$; there are only finitely many such groups. □

Proof of Theorem 3.3. Let $\check{\mu}$ denote the image measure of μ by the map $g \mapsto g^{-1}$. By lemmata 3.4 and 3.5 above, there is $k \geq 1$ such that the support of $\check{\mu}^{*k} * \mu^{*k}$ contains a finite symmetric subset S which generates a Zariski dense subgroup in $\mathbb{G}$. By [38, Theorem 1] applied to S, there is $c > 0$ such that for any prime number p sufficiently large,

$$\|T_{\mu_S}\|_{\ell^2_0(\mathbb{G}_p(\mathbb{F}_p))} \leq 1 - c,$$

where μ_S denotes the normalized counting measure on S. Then, we can write

$$\check{\mu}^{*k} * \mu^{*k} = \alpha\mu_S + (1-\alpha)\mu'$$

where $\alpha > 0$ and μ' is some probability measure on $\mathrm{SL}_d(\mathbb{Z})$. Thus, for any prime number p sufficiently large, using the fact that $T_\mu^* = T_{\check{\mu}}$,

$$\begin{aligned}
\|T_\mu^k\|^2_{\ell_0^2(\mathbb{G}_p(\mathbb{F}_p))} &\leq \|(T_\mu^*)^k T_\mu^k\|_{\ell_0^2(\mathbb{G}_p(\mathbb{F}_p))} \\
&= \|T_{\check{\mu}^{*k} * \mu^{*k}}\|_{\ell_0^2(\mathbb{G}_p(\mathbb{F}_p))} \\
&\leq \alpha \|T_{\mu_S}\|_{\ell_0^2(\mathbb{G}_p(\mathbb{F}_p))} + (1-\alpha)\|T_{\mu'}\|_{\ell_0^2(\mathbb{G}_p(\mathbb{F}_p))} \\
&\leq 1 - \alpha c.
\end{aligned}$$

□

One can also interpret Theorem 3.3 as a statement on the speed of equidistribution of the random walks associated to μ on the Cayley graphs $\mathbb{G}_p(\mathbb{F}_p)$. This is explained for instance in [30, §3.1] for the case of simple random walks on a family of expander graphs. In our setting, we obtain the following corollary, whose proof is left to the reader. For a prime number p, we denote by $\pi_p\colon \mathbb{Z} \to \mathbb{F}_p$ the reduction modulo p. By abuse of notation we extend the domain of definition of π_p to any free $\mathbb{Z}$-module.

Corollary 3.6. *Let $d \geq 2$ and let μ be a probability measure on $\mathrm{SL}_d(\mathbb{Z})$ such that the Zariski closed subgroup $\mathbb{G}$ generated by μ is semisimple. There exists $C \geq 0$ such that for every prime number p sufficiently large, and $n \geq C \log p$, for all $a \in \mathcal{M}_d(\mathbb{F}_p)$,*

$$\mu^{*n}(\{g \in \mathcal{M}_d(\mathbb{Z}) \mid \pi_p(g) = a\}) \leq \frac{2}{|\mathbb{G}_p(\mathbb{F}_p)|}.$$

We conclude this paragraph by a lemma that will allow us to use the spectral gap theorem in our setting of random walks on the torus.

Lemma 3.7 (Strong irreducibility and semisimplicity)**.** *Let G be a Zariski closed subgroup of $\mathrm{SL}_d(\mathbb{R})$ generated by elements of $\mathrm{SL}_d(\mathbb{Z})$ and acting strongly irreducibly on $\mathbb{R}^d$. Then G is semisimple.*

Proof. Since G acts irreducibly on a finite-dimensional vector space, it is a reductive group, and can be written as an almost product

$$G = Z \cdot S,$$

where Z is a torus, central in G, and S is semisimple, with $Z \cap S$ finite. The group Z is equal to the intersection of G with the center of the algebra E generated by G. Since E is a simple algebra over $\mathbb{R}$, its center can be identified with $\mathbb{R}$ or $\mathbb{C}$. Note that the restriction of the determinant on $\mathcal{M}_d(\mathbb{R})$ to the center of E is simply a power of the usual norm on $\mathbb{C}$, and since $G \subset \mathrm{SL}_d(\mathbb{R})$, the group Z must be included in the group of complex numbers of norm 1. In particular, $G/S \simeq Z/Z \cap S$ is compact.

Now since G is defined over $\mathbb{Q}$, the projection map $G \to G/S$ is given by some polynomial map with rational coefficients. In particular, the image F of $G \cap \mathrm{SL}_d(\mathbb{Z})$ inside $Z/Z \cap S$ is made of matrices whose entries are rational with bounded denominators, and is therefore finite. But G equals the Zariski closure of $G \cap \mathrm{SL}_d(\mathbb{Z})$, so the image of G itself under the projection $G \to G/S$ is finite, i.e. G is semisimple. □

3.2. **Escaping from subvarieties: a consequence of the spectral gap.** The aim of this subsection is to establish the following proposition, using the results of the previous paragraph.

Proposition 3.8. *Let μ be a probability measure on $\mathrm{SL}_d(\mathbb{Z})$ such that the Zariski closed subgroup $\mathbb{G}$ generated by* $\mathrm{Supp}\,\mu$ *is semisimple and connected.*

There exists a constant $c > 0$ depending on μ such that for all polynomials $f \in \mathbb{Q}[\mathcal{M}_d]$, of degree at most D and not vanishing on $\mathbb{G}$, we have

$$\forall n \geq 1, \quad \mu^{*n}(\{g \in \Gamma \mid f(g) = 0\}) \ll_{\mu,D} e^{-cn}.$$

Indeed, this is a general version of [13, Corollary 1.1] which is stated for the group SL_d and for the simple random walk on the Cayley graph. The proof is essentially the same. But since it will be important for us to know that the upper bound depends only on the degree of f, we provide a detailed argument.

We need the following lemma, which is a consequence of the Lang-Weil inequality. It will be important for us to have an estimate which is uniform for subvarieties of bounded complexity.

Lemma 3.9. *Given a geometrically irreducible subvariety $V \subset \mathbb{A}^d$ defined over $\mathbb{Q}$ and an integer $D \geq 1$, there exists $p_0 = p_0(V, D)$ such that the following holds. Let $f \in \mathbb{Q}[X_1, \ldots, X_d]$ be a polynomial of degree at most D. Assume that f does not vanish on V. Then for every prime number $p \geq p_0$,*

$$|\{\pi_p(x) \in \mathbb{F}_p^d \mid x \in V \cap \mathbb{Z}^d,\, f(x) = 0\}| \ll_{V,D} p^{-1}|V_p(\mathbb{F}_p)|$$

where V_p denotes the reduction modulo p of V.

Note that to speak about reductions modulo p of a variety over $\mathbb{Q}$, we need to choose a model over $\mathbb{Z}$. But since V is a subvariety of an affine space, among such choices, there is a canonical one. See the first paragraph in the proof below.

Proof. We abbreviate $\mathbb{Q}[X_1, \ldots, X_d]$ as $\mathbb{Q}[X]$ and $\mathbb{Z}[X_1, \ldots, X_d]$ as $\mathbb{Z}[X]$. Let $I \subset \mathbb{Q}[X]$ be the ideal of all polynomials with coefficients in $\mathbb{Q}$ and vanishing on V. Let $I_{\mathbb{Z}} = I \cap \mathbb{Z}[X]$. For any prime number p, let V_p be the variety over $\mathbb{F}_p$ defined by the ideal $\pi_p(I_{\mathbb{Z}}) \subset \mathbb{F}_p[X]$. By the Bertini-Noether theorem [21, Proposition 10.4.2], V_p is geometrically irreducible for $p \geq p_0$, where p_0 is a constant depending only on the embedding $V \subset \mathbb{A}^d$.

On the one hand applying the Lang-Weil inequality [31, Theorem 1] to the irreducible variety V_p, we obtain

$$|V_p(\mathbb{F}_p)| \geq \frac{1}{2} p^{\dim(V_p)}. \tag{3.1}$$

On the other hand, one can prove, using Gröbner bases [19, Chapter 2], that given an integer $D \geq 1$ there is $p_0 = p_0(V, D)$ such that for all $p \geq p_0$ and all $f \in \mathbb{Q}[X] \setminus I$ of degree at most D there is $h \in (\mathbb{Q}f \oplus I) \cap \mathbb{Z}[X]$ of degree at most $O_{V,D}(1)$ and such that $\pi_p(h) \notin \pi_p(I_{\mathbb{Z}})$. For such h, we have

$$\{x \in V \cap \mathbb{Z}^d \mid f(x) = 0\} \subset \{x \in V \cap \mathbb{Z}^d \mid h(x) = 0\}$$

and hence

$$|\{\pi_p(x) \in \mathbb{F}_p^d \mid x \in V \cap \mathbb{Z}^d,\, f(x) = 0\}| \leq |\{x \in V_p(\mathbb{F}_p) \mid \pi_p(h)(x) = 0\}|.$$

The right-hand side is the number of $\mathbb{F}_p$-points in the subvariety $V_p \cap \{\pi_p(h) = 0\}$. This subvariety has dimension at most $\dim(V_p) - 1$ since V_p is irreducible and $\pi_p(h) \notin \pi_p(I_{\mathbb{Z}})$. Thus, applying a version of the Schwarz-Zippel estimate, like [31, Lemma 1], and using the fact that the complexity controls the degree, we get

$$|(V_p \cap \{\pi_p(h) = 0\})(\mathbb{F}_p)| \ll_{V,D} p^{\dim(V_p)-1}. \tag{3.2}$$

Together with (3.1), this proves the desired inequality. $\square$

Now we are ready to prove Proposition 3.8.

Proof of Proposition 3.8. By Lemma 3.9,

$$|\{\pi_p(g) \in \mathcal{M}_d(\mathbb{F}_p) \mid g \in \Gamma,\, f(g) = 0\}| \ll_{\mathbb{G},D} p^{-1}|\mathbb{G}_p(\mathbb{F}_p)|$$

for every prime number $p \geq p_0(\mathbb{G}, D)$. Combined with Corollary 3.6, this yields

$$\mu^{*n}(\{g \in \Gamma \mid f(g) = 0\}) \ll_{\mathbb{G},D} p^{-1}.$$

for all $n \geq C \log p$, where C is a constant depending only on μ. We conclude by choosing p to be a prime number such that $p \asymp e^{n/C}$ and $p \geq p_0$. $\square$

3.3. **Interlude : large deviation estimates for random matrix products.** In this subsection, μ is a Borel probability measure on $\mathrm{SL}_d(\mathbb{R})$, not necessarily supported on matrices with integer coefficients. By Γ we denote the closure of the subsemigroup generated by $\mathrm{Supp}(\mu)$ and by G the group of $\mathbb{R}$-points of the Zariski closure of Γ.

Let us first recall the large deviation estimates for random matrix products. This result is originally due to Lepage [32], and the version below is taken from Bougerol [8, Theorem V.6.2]. For $g \in \mathrm{GL}_d(\mathbb{R})$, denote by $\sigma_1(g) \geq \cdots \geq \sigma_d(g) > 0$ the singular values of g ordered decreasingly.

Theorem 3.10 (Large deviation estimates)**.** *Let μ be a Borel probability measure on $\mathrm{GL}_d(\mathbb{R})$ having a finite exponential moment. Let λ_k denote the k-th Lyapunov exponent associated to μ. Assume that G acts strongly irreducibly on $\mathbb{R}^d$. For any $\omega > 0$, there is $c > 0$, $n_0 > 0$ such that the following holds.*

(i) For all $n \geq n_0$,

$$\mu^{*n}\big(\{g \in \Gamma \mid |\frac{1}{n}\log\|g\| - \lambda_1| \geq \omega\}\big) \leq e^{-cn}.$$

(ii) For all $k = 1, \dots, d$ and all $n \geq n_0$,

$$\mu^{*n}\big(\{g \in \Gamma \mid |\frac{1}{n}\log \sigma_k(g) - \lambda_k| \geq \omega\}\big) \leq e^{-cn}.$$

(iii) For all $n \geq n_0$, For all $v \in \mathbb{R}^d \setminus \{0\}$,

$$\mu^{*n}\big(\{g \in \Gamma \mid \big|\frac{1}{n}\log\frac{\|gv\|}{\|v\|} - \lambda_1\big| \geq \omega\}\big) \leq e^{-cn}.$$

Item (i) is, of course, a special case of Item (ii) since $\sigma_1(g) = \|g\|$. One consequence of the above theorem that will be useful to us is the following proposition.

Proposition 3.11. *Let μ be a Borel probability measure on $\mathrm{GL}_d(\mathbb{R})$ having a finite exponential moment. Assume that G acts strongly irreducibly on $\mathbb{R}^d$. Then there exists $\kappa > 0$ such that for all $n \geq 1$ and all $\rho \geq e^{-n}$, for every $v \in \mathbb{R}^d \setminus \{0\}$,*

$$\mu^{*n}\big(\{g \in \Gamma \mid \|gv\| \leq \rho\|g\|\|v\|\}\big) \ll_\mu \rho^\kappa.$$

Proof. Let r denote the proximal dimension of G, i.e.

$$r = \min\{\mathrm{rank}\, g \ ;\ g \in \overline{\mathbb{R}G} \setminus \{0\}\}.$$

For $g \in G$, write its Cartan decomposition $g = k\,\mathrm{diag}(\sigma_1(g), \dots, \sigma_d(g))\ell$, where k and ℓ are orthogonal matrices and $\sigma_1(g) \geq \cdots \geq \sigma_d(g)$ are the singular values of g. Define also

$$V_g^+ = k\,\mathrm{Span}(e_1, \dots, e_r) \ \text{ and } \ W_g^- = \ell^{-1}\,\mathrm{Span}(e_{r+1}, \dots, e_d)$$

where $(e_1, \dots, e_d)$ is the standard basis of $\mathbb{R}^d$.

We first prove the proposition in the case where G is proximal, i.e. $r = 1$. Under this condition, we have [11, Lemma 4.1(2)] :

$$\forall g \in \mathrm{GL}_d(\mathbb{R}), \forall v \in \mathbb{R}^d \setminus \{0\}, \quad \|gv\| \geq \mathrm{d}_\measuredangle(\mathbb{R}v, W_g^-)\|g\|\|v\|, \tag{3.3}$$

where $\mathrm{d}_\measuredangle$ is defined, for any two subspaces $V, W \subset \mathbb{R}^d$ of $\mathbb{R}^d$ with respective orthonormal bases $(v_1, \dots, v_s)$ and $(w_1, \dots, w_t)$, by the formula

$$\mathrm{d}_\measuredangle(V, W) = \|v_1 \wedge \cdots \wedge v_s \wedge w_1 \wedge \cdots \wedge w_t\|.$$

Then, [11, Lemma 4.5] applied to the transposed random walk, which is also proximal, shows that there exists $\kappa > 0$ such that

$$\forall \rho \geq e^{-n}, \quad \mu^{*n}\big(\{g \in \Gamma \mid \mathrm{d}_\measuredangle(V^+_{{}^tg}, v^\perp) \leq \rho\}\big) \ll_\mu \rho^\kappa. \tag{3.4}$$

Noting that $(W_g^-)^\perp = V^+_{{}^tg}$ and that $\mathrm{d}_\measuredangle(V, W) = \mathrm{d}_\measuredangle(V^\perp, W^\perp)$ if $\dim(V)+\dim(W) = d$, we see that the result follows from (3.3).

We now use Bougerol's trick [8, Proof of Theorem V.6.2] to reduce the proposition to the proximal case. Namely, by [17, Lemma 3.2] or [6, Lemma 4.36], we have a decomposition of $\wedge^r\mathbb{R}^d$ into G-invariant subspaces

$$\wedge^r\mathbb{R}^d = \Lambda_+ \oplus \Lambda_0$$

such that the action of G on Λ_+ is strongly irreducible and proximal and moreover,

$$\forall g \in G, \quad \|(\wedge^r g)_{|\Lambda_+}\| \gg_G \|g\|^r \geq \|\wedge^r g\|. \tag{3.5}$$

Denote by $\pi_+ \colon \wedge^r\mathbb{R}^d \to \Lambda_+$ the projection with respect to this decomposition. By [17, Lemma 3.3], for any $v \in \mathbb{R}^d$, we can find a subspace $P \subset \mathbb{R}^d$ of dimension r and containing v such that

$$\pi_+(\mathbf{v}_P) \gg_G \|\mathbf{v}_P\|, \tag{3.6}$$

where $\mathbf{v}_P \in \wedge^r\mathbb{R}^d$ is the wedge product of the elements of a basis of P. Now, observe that (3.3) still holds without the proximality assumption, and we have, for every g in $\mathrm{GL}_d(\mathbb{R})$,

$$\mathrm{d}_\measuredangle(\mathbb{R}v, W_g^-) \geq \mathrm{d}_\measuredangle(P, W_g^-).$$

Hence, it suffices to prove, for some $\kappa > 0$ and $n_0 \geq 1$,

$$\forall n \geq n_0, \forall \rho \geq e^{-n}, \quad \mu^{*n}\big(\{g \in \Gamma \mid \mathrm{d}_\measuredangle(P, W_g^-) \leq \rho\}\big) \ll_\mu \rho^\kappa. \tag{3.7}$$

By [17, Lemma 4.2],

$$\forall g \in \mathrm{GL}_d(\mathbb{R}), \quad \frac{\|(\wedge^r g)\mathbf{v}_P\|}{\|\wedge^r g\|\,\|\mathbf{v}_P\|} \leq \mathrm{d}_\measuredangle(P, W_g^-) + \frac{\sigma_{r+1}(g)}{\sigma_r(g)}.$$

Combined with (3.5) and (3.6), this yields

$$\forall g \in G, \quad \frac{\|(\wedge^r g)_{|\Lambda_+}\pi_+(\mathbf{v}_P)\|}{\|(\wedge^r g)_{|\Lambda_+}\|\,\|\pi_+(\mathbf{v}_P)\|} \ll_G \mathrm{d}_\measuredangle(P, W_g^-) + \frac{\sigma_{r+1}(g)}{\sigma_r(g)}.$$

By a result of Guivarc'h-Raugi [23], $\lambda_r > \lambda_{r+1}$. Applying Theorem 3.10(ii) to $k = r$ and $r+1$ and $\omega_0 = (\lambda_r - \lambda_{r+1})/3 > 0$, we get $c > 0$ and $n_0 \geq 1$ such that

$$\forall n \geq n_0, \quad \mu^{*n}\Big(\Big\{g \in \Gamma \mid \frac{\sigma_{r+1}(g)}{\sigma_r(g)} \leq e^{-\omega_0 n}\Big\}\Big) \geq 1 - e^{-cn}.$$

Note that $e^{-cn} \leq \rho^c$. The desired estimate (3.7) then follows by the proximal case applied to the induced random walk on Λ_+ and to the vector $\pi_+(\mathbf{v}_P) \in \Lambda_+$. ☐

3.4. **Escaping a small neighborhood of a subvariety.** For an integer $D \geq 0$, and a regular function $f \in \mathbb{R}[\mathbb{G}]$ on $\mathbb{G}$, we say that f has degree at most D if it can be represented by a polynomial on $\mathcal{M}_d$ of degree at most D. Denote by $\mathbb{R}[\mathbb{G}]_{\leq D}$ the finite-dimensional subspace consisting of regular functions of degree at most D. We fix a norm on $\mathbb{R}[\mathbb{G}]_{\leq D}$.

Lemma 3.12. *Let μ be a Borel probability measure on* $\mathrm{SL}_d(\mathbb{Z})$ *having a finite exponential moment. Assume that the Zariski closed subgroup $\mathbb{G}$ generated by* $\operatorname{Supp}\mu$ *is semisimple and connected.*

Given an integer $D \geq 1$, there exist constants $C > 0$, $c > 0$ and $n_0 \geq 1$ depending on μ and D such that

$$\forall f \in \mathbb{R}[\mathbb{G}]_{\leq D},\ \forall n \geq n_0, \quad \mu^{*n}\big(\{g \in \Gamma \mid |f(g)| < e^{-Cn}\|f\|\}\big) \leq e^{-cn}.$$

Proof. By Theorem 3.10(i), there is $c > 0$ such that for n large enough

$$\mu^{*n}\big(\{g \in \Gamma \mid \|g\| \geq e^{2\lambda_1 n}\}\big) \leq e^{-cn},$$

where λ_1 is the top Lyapunov exponent associated to the random walk defined by μ on $\mathbb{R}^d$. Thus, we are left to bound from above the μ^{*n}-measure of the set

$$A_C = \big\{g \in \Gamma \mid |f(g)| \leq e^{-Cn}\|f\| \text{ and } \|g\| \leq e^{2\lambda_1 n}\big\}.$$

Let us abbreviate $V = \mathbb{R}[\mathbb{G}]_{\leq D}$. Let $V_{\mathbb{Q}}$ denote the set of functions $f \in V$ which are $\mathbb{Q}$-rational, i.e. represented by polynomials on $\mathcal{M}_d$ with coefficients in $\mathbb{Q}$. Note that $V_{\mathbb{Q}}$ defines a $\mathbb{Q}$-structure on V.

We claim that if C is chosen large enough then A_C must be contained in some subvariety $\{f_0 = 0\}$ where $f_0 \in V_{\mathbb{Q}} \setminus \{0\}$. Then Proposition 3.8 applied to f_0 allows to conclude.

For every $g \in \Gamma$ let $\mathrm{ev}_g : V \to \mathbb{R}$ be the evaluation map

$$\forall v \in V, \quad \mathrm{ev}_g(v) = v(g).$$

Since the matrices $g \in A_C$ have integer coefficients, the intersection

$$W = \bigcap_{g \in A_C} \ker(\mathrm{ev}_g)$$

is a subspace of V defined over $\mathbb{Q}$, i.e. $W = \mathbb{R} \otimes_{\mathbb{Q}} (W \cap V_{\mathbb{Q}})$. We want to show that $W \cap V_{\mathbb{Q}}$ contains nonzero element. Assume for a contradiction that $W = \{0\}$.

Write $s = \dim(V)$. We can choose $g_1, \dots, g_s \in A_C$ such that

$$\{0\} = \bigcap_{i=1}^{s} \ker(\mathrm{ev}_{g_i}).$$

Fix a basis $(v_1, \dots, v_s)$ of V in which each element is represented by a polynomial on $\mathcal{M}_d$ with coefficients in $\mathbb{Z}$. Thus, the map $\Phi \colon V \to \mathbb{R}^s$ defined by

$$\forall f \in V, \quad \Phi(v) = \big(\mathrm{ev}_{g_i}(v)\big)_i$$

is invertible and has integer coefficients when expressed in the basis $(v_1, \dots, v_s)$ and the standard basis of $\mathbb{R}^s$. Thus, in these bases, the determinant of Φ satisfies $|\det(\Phi)| \geq 1$. Moreover,

$$\|\Phi\| \ll \max_{1 \leq i,j \leq s} |\mathrm{ev}_{g_i}(v_j)| \ll_{\mathbb{G},D} \max_{1 \leq i \leq s} \|g_i\|^D \leq e^{2D\lambda_1 n}.$$

It follows that

$$\|\Phi^{-1}\| \ll \frac{\|\Phi\|^{s-1}}{|\det(\Phi)|} \leq e^{2Ds\lambda_1 n}.$$

For $f \in V$, by definition of A_C, we have

$$\|\Phi(f)\| = \max_{1 \leq i \leq s} |f(g_i)| \leq e^{-Cn}\|f\|.$$

Thus,

$$\|f\| \leq \|\Phi^{-1}\|\|\Phi(f)\| \ll_{\mathbb{G},D} e^{(2Ds\lambda_1 - C)n}\|f\|.$$

We get a contradiction if C is chosen to be larger than $2Ds\lambda_1 + O_{\mathbb{G},D}(1)$. □

The following is a variant and an easy consequence of the previous lemma.

Lemma 3.13. *Let μ be a Borel probability measure on $\mathrm{SL}_d(\mathbb{Z})$ having a finite exponential moment. Assume that the Zariski closed subgroup $\mathbb{G}$ generated by $\operatorname{Supp}\mu$ is semisimple and connected.*

Given an integer D, there exist constants $C > 0$, $c > 0$ and $n_0 \geq 1$ depending on μ and D such that

$$\forall f \in \mathbb{R}[\mathbb{G}]_{\leq D},\ \forall n \geq n_0, \quad \mu^{*n}\big(\{g \in \Gamma \mid |f(g)| < e^{-Cn}\|g\|\|f\|\}\big) \leq e^{-cn}.$$

Proof. For any $n \geq 1$ and any $C > 0$, we have

$$\begin{aligned}\mu^{*n}\big(\{g \in \Gamma \mid |f(g)| &< e^{-(C+2\lambda_1)n}\|g\|\|f\|\}\big)\\ &\leq \mu^{*n}\big(\{g \in \Gamma \mid |f(g)| < e^{-Cn}\|f\|\}\big) + \mu^{*n}\big(\{g \in \Gamma \mid \|g\| \geq e^{2\lambda_1 n}\}\big).\end{aligned}$$

We conclude by using Lemma 3.12 for the first term and Theorem 3.10(i) for the second. □

3.5. **Non-concentration near affine subspaces.** We now want to prove Proposition 3.1. Of course, if we are to show that the random walk does not concentrate near any proper affine subspace in the algebra E, we should first check that the group G is not trapped in any proper affine subspace.

Lemma 3.14. *Let G be a subgroup of $\mathrm{GL}_d(\mathbb{R})$ acting irreducibly on $\mathbb{R}^d$, and E the associative subalgebra generated by G in $\mathcal{M}_d(\mathbb{R})$. Then G is not contained in any proper affine subspace of E.*

Proof. Equivalently, we have to show that the linear span $W = \operatorname{Span}(G - 1)$ of $G-1$ is E. For this, it suffices to prove that $1 \in W$. Firstly, W is closed under multiplication. Indeed, any product between two elements of W is a linear combination of elements of the form $(g-1)(h-1)$ with $g, h \in G$ and we have

$$(g-1)(h-1) = (gh-1) - (g-1) - (h-1) \in W.$$

Secondly, any subspace of $\mathbb{R}^d$ preserved by W is preserved by G, hence the only subspaces preserved by W are $\mathbb{R}^d$ and $\{0\}$. We conclude by using the following algebraic lemma. □

Momentarily, in the next lemma and its proof, algebras are not assumed to be unital. Thus, an subalgebra is a linear subspace that is closed under multiplication. Accordingly, a left (resp. right) ideal, is a subspace preserved under multiplication on the left (resp. right) by all elements of the algebra.

Lemma 3.15. *If a nonzero subalgebra of $\mathcal{M}_d(\mathbb{R})$ does not preserve any proper nontrivial subspaces of $\mathbb{R}^d$, then it contains the multiplicative identity of $\mathcal{M}_d(\mathbb{R})$.*

Proof. Let $W \subset \mathcal{M}_d(\mathbb{R})$ be such a subalgebra. We first show that the only nilpotent right ideal of W is the zero ideal. Indeed, let I be a nonzero nilpotent right ideal of W. Let k be the maximal number such that $I^k \neq 0$. Let $f_0 \in I^k$. Since $I^{k+1} = 0$, we have

$$f_0(\mathbb{R}^d) \subset \bigcap_{f \in I} \ker(f).$$

The intersection on the right-hand side is preserved by W and not equal to $\mathbb{R}^d$, because I is a nonzero right ideal. Then f_0 must be zero, which contradicts $I^k \neq 0$.

Thus, W is an algebra without radical [41, Chapter XVI, §116]. By [41, Chapter XVI, §117][2], W has a multiplicative identity 1_W. Its image $1_W(\mathbb{R}^d)$ is preserved by W and nonzero since W is nonzero. Hence $1_W(\mathbb{R}^d) = \mathbb{R}^d$ and $1_W \in \mathrm{GL}_d(\mathbb{R})$. Then $1_W^2 = 1_W$ forces 1_W to be the identity of $\mathcal{M}_d(\mathbb{R})$. □

[2]More precisely, we apply the theorem to algebras, or rings with operator domain $\mathbb{R}$, using the terminology of Van der Waerden, see [41, Chapter XVI, §115].

Now we are ready to prove Proposition 3.1.

Proof of Proposition 3.1. By Lemma 3.14, $\mathbb{R}[\mathbb{G}]_{\leq 1}$ is isomorphic to $\mathbb{R} \oplus E^*$, the space of affine mappings from E to $\mathbb{R}$. On $\mathbb{R} \oplus E^*$, let G act by

$$\forall g \in G,\ \forall f \in \mathbb{R} \oplus E^*,\ \forall x \in E,\ (g \cdot f)(x) = f(xg).$$

By Lemma 3.13, there exist $C_1 \geq 1$ and $c_1 > 0$ such that

$$\forall f \in \mathbb{R} \oplus E^*,\ \forall m \geq 1, \quad \mu^{*m}\big(\{g \in \Gamma \mid |f(g)| < e^{-C_1 m}\|g\|\|f\|\}\big) \ll_\mu e^{-c_1 m}. \tag{3.8}$$

Given a proper affine subspace $W \subset E$, there exists $f \in \mathbb{R} \oplus E^*$ such that its linear part $f_1 \in E^*$ has norm $\|f_1\| = 1$ and

$$\forall g \in E,\ d(g, W) = |f(g)|.$$

Let $\rho \geq e^{-n}$. Pick m such that $e^{-C_1 m} \asymp \rho^{1/2}$. Using the relation $\mu^{*n} = \mu^{*m} * \mu^{*(n-m)}$, we have

$$\mu^{*n}(\{g \in \Gamma \mid |f(g)| \leq \rho\|g\|\}) = \int_\Gamma \mu^{*m}(\{g \in \Gamma \mid |(h \cdot f)(g)| \leq \rho\|gh\|\}) \,\mathrm{d}\mu^{*(n-m)}(h)$$

We distinguish two cases according to whether $\rho\|h\| \leq e^{-C_1 m}\|h \cdot f\|$. If this is the case, then $\rho\|gh\| \leq e^{-C_1 m}\|g\|\|h \cdot f\|$ and then by (3.8),

$$\mu^{*m}\big(\{g \in \Gamma \mid |(h \cdot f)(g)| \leq \rho\|gh\|\}\big) \ll_\mu e^{-c_1 m} \ll \rho^{c_1/2C_1}.$$

Otherwise, $\|h \cdot f\| \leq \rho e^{C_1 m}\|h\| \leq \rho^{1/2}\|h\|$ by the choice of m.

Thus,

$$\mu^{*n}(\{g \in \Gamma \mid |f(g)| \leq \rho\|g\|\}) \ll_\mu \rho^{c_1/2C_1} + \mu^{*(n-m)}\big(\{h \in \Gamma \mid \|h \cdot f\| \leq \rho^{1/2}\|h\|\}\big).$$

Now observe that E^* is a submodule of the the semisimple G-module $\mathcal{M}_d(\mathbb{R})^*$, which is isomorphic to the sum of d copies of the simple G-module $\mathbb{R}^d$. It follows that E^* is isomorphic to the sum of $\frac{\dim(E)}{d}$ copies of $\mathbb{R}^d$. For each $i = 1, \ldots, \frac{\dim(E)}{d}$, let $\pi_i \colon \mathbb{R} \oplus E^* \to \mathbb{R}^d$ denote the projection to the i-th $\mathbb{R}^d$-factor. Remembering $\|f_1\| = 1$, we obtain $\|\pi_i(f)\| \gg_G 1$ for some i. Then $\|h \cdot f\| \leq \rho^{1/2}\|h\|$ implies $\|h\pi_i(f)\| \ll_G \rho^{1/2}\|h\|\|\pi_i(f)\|$. Hence

$$\begin{aligned}\mu^{*(n-m)}&\big(\{h \in \Gamma \mid \|h \cdot f\| \leq \rho^{1/2}\|h\|\}\big)\\ &\leq \mu^{*(n-m)}\big(\{h \in \Gamma \mid \|h\pi_i(f)\| \leq C_2\rho^{1/2}\|h\|\|\pi_i(f)\|\}\big)\end{aligned}$$

where C_2 is a constant depending only on G.

By our choice of m, we have $C_2\rho^{1/2} \geq e^{-(n-m)}$. Hence, by Proposition 3.11,

$$\mu^{*(n-m)}\big(\{h \in \Gamma \mid \|h\pi_i(f)\| \leq C_2\rho^{1/2}\|h\|\|\pi_i(f)\|\}\big) \ll_\mu \rho^{\kappa_2/2}$$

where $\kappa_2 > 0$ is the constant given by Proposition 3.11 which depends only on μ. This proves the desired estimate with $\kappa = \min\{\frac{c_1}{2C_1}, \frac{\kappa_2}{2}\}$. ☐

3.6. **Escaping a larger neighborhood of a subvariety.** The rest of this section is devoted to the proof of Proposition 3.2. The idea is to generalize what we did above for affine subspaces. This time, the variety that we want to avoid is defined by a general polynomial map f on the algebra E, so that we shall have to consider the representation $\rho \colon G \to \mathrm{GL}(\mathbb{R}[\mathbb{G}])$ defined by

$$\forall g \in G,\ \forall f \in \mathbb{R}[\mathbb{G}],\ \forall x \in \mathbb{G}, \quad (\rho(g)f)(x) = f(xg).$$

We refer to finite-dimensional subrepresentations of this representation as G-modules. For a G-module M, we denote by $\lambda_1(\mu, M)$ the top Lyapunov exponent associated to the random walk on M defined by μ:

$$\lambda_1(\mu, M) = \lim \frac{1}{n}\int_G \log\|\rho(g)_{|M}\| \,\mathrm{d}\mu^{*n}(g)$$

where $\| \ \|$ denotes some operator norm.

For a real number $\lambda \geq 0$, define M_λ to be the sum of submodules M of $\mathbb{R}[\mathbb{G}]_{\leq D}$ such that $\lambda_1(\mu, M) \geq \lambda$. Let $p_\lambda \colon \mathbb{R}[\mathbb{G}]_{\leq D} \to M_\lambda$ be an epimorphism of G-modules onto M_λ. Remark that M_λ is a sum of isotypical components in $\mathbb{R}[\mathbb{G}]_{\leq D}$ so that p_λ is uniquely defined.

Proposition 3.16. *Let μ be a probability measure on $\mathrm{SL}_d(\mathbb{Z})$, for some $d \geq 2$. Let Γ denote the subsemigroup generated by μ, $\mathbb{G}$ the Zariski closure of Γ in SL_d, and λ_1 the top Lyapunov exponent of μ. Assume that:*

(a) The measure μ has a finite exponential moment;
(b) The action of Γ on $\mathbb{R}^d$ is irreducible;
(c) The algebraic group $\mathbb{G}$ is Zariski connected,

and let the notation be as above.

Given D, $\lambda \geq 0$ and $\omega > 0$, there is $c > 0$ and $n_0 \geq 1$ (depending also on μ) such that

$$\forall f \in \mathbb{R}[\mathbb{G}]_{\leq D},\ \forall n \geq n_0, \quad \mu^{*n}\big(\{g \in \Gamma \mid |f(g)| \leq e^{(\lambda-\omega)n}\|p_\lambda(f)\|\}\big) \leq e^{-cn}.$$

Proof. Note that, for any $0 < m < n$, we have

$$\begin{aligned} \mu^{*n}\big\{g \in \Gamma \mid |f(g)| \leq e^{(\lambda-\omega)n}\|p_\lambda(f)\|\big\} \\ = \int_\Gamma \mu^{*m}\big\{g \in \Gamma \mid |(\rho(h)f)(g)| \leq e^{(\lambda-\omega)n}\|p_\lambda(f)\|\big\}\, \mathrm{d}\mu^{*(n-m)}(h) \end{aligned} \tag{3.9}$$

For any $f \in \mathbb{R}[\mathbb{G}]_{\leq D}$, there is a simple G-module M contained in M_λ such that

$$\|p_M(f)\| \gg_{\mathbb{G},D,\lambda} \|p_\lambda(f)\|$$

where $p_M \colon M_\lambda \to M$ is a projection of G-modules. By definition of M_λ, the top Lyapunov exponent in M satisfies $\lambda_1(\mu, M) \geq \lambda$. Note that since $\mathbb{G}$ is Zariski connected, the G-action on the simple module M is strongly irreducible. Hence, by the large deviation estimate Theorem 3.10(iii), there is $c > 0$ and $n_0 \geq 1$ such that for all $m > 0$ satisfying $n - m \geq n_0$,

$$\mu^{*(n-m)}\big\{h \in \Gamma \mid \|\rho(h)p_M(f)\| \leq e^{(\lambda-\frac{\omega}{2})(n-m)}\|p_M(f)\|\big\} \leq e^{-c(n-m)},$$

and hence

$$\mu^{*(n-m)}\big\{h \in \Gamma \mid \|\rho(h)f\| \leq e^{(\lambda-\frac{\omega}{2})(n-m)}\|p_\lambda(f)\|\big\} \leq e^{-c(n-m)}. \tag{3.10}$$

Applying Lemma 3.12 to the function $\rho(h)f$, for $h \in \Gamma$, we obtain $\forall m \geq m_0$

$$\mu^{*m}\big\{g \in \Gamma \mid |(\rho(h)f)(g)| < e^{-Cm}\|\rho(h)f\|\big\} \leq e^{-cm}, \tag{3.11}$$

for some $C > 0$, $c > 0$ and $m_0 \geq 1$. Setting $m = \big\lfloor \frac{\omega n}{2(C+\lambda)} \big\rfloor$, so that

$$(\lambda - \omega)n - (\lambda - \frac{\omega}{2})(n-m) \leq -Cm,$$

the desired inequality follows from (3.9), (3.10) and (3.11). $\square$

3.7. **Criterion to have nonzero component in modules of maximal Lyapunov exponent.** In order to use Proposition 3.16, we need to be able to say when a regular function has a nonzero component in a simple submodule of large Lyapunov exponent.

Lemma 3.17. *Let μ be a probability measure on $\mathrm{SL}_d(\mathbb{R})$, $d \geq 2$, with some finite exponential moment. Assume that the group Γ generated by μ is non-compact and acts irreducibly on $\mathbb{R}^d$, and that its Zariski closure $\mathbb{G}$ is connected.*

Let $f \in \mathbb{R}[\mathcal{M}_d]$ be a polynomial of degree $D \geq 1$ whose degree D homogeneous part does not vanish on the algebra E generated by G. The following holds for every integer $k \geq \dim(E)$. Consider the polynomial $\bar{F} \in \mathbb{R}[\mathcal{M}_d^k]$ defined by

$$\bar{F}(x_1, \dots, x_k) = f(x_1 + \dots + x_k)$$

and let $F \in \mathbb{R}[\mathbb{G}^k]$ be the restriction of $\bar{F}$ to $\mathbb{G}^k$. Then

$$p(F) \neq 0$$

where $p \colon \mathbb{R}[\mathbb{G}^k]_{\leq D} \to \mathbb{R}[\mathbb{G}^k]_{\leq D}$ is the projection to the sum of all simple G^k-submodules M of $\mathbb{R}[\mathbb{G}^k]_{\leq D}$ having $\lambda_1(\mu^{\otimes k}, M) \geq D\lambda_1(\mu, \mathbb{R}^d)$.

We shall use the theory of the highest weight as well as the theory of random walks on semisimple groups. So let us fix some notation and recall briefly the needed results. Let $\mathfrak{g}$ denote the Lie algebra of G. Let K be a maximal compact subgroup of G. Inside the orthogonal complement, with respect to the Killing form, of the Lie algebra of K, we choose a Cartan subspace $\mathfrak{a}$ of $\mathfrak{g}$. Every algebraic representation of G is diagonalizable for $\mathfrak{a}$. That is, for every G-module M, we have

$$M = \bigoplus_{\chi \in \mathfrak{a}} M_\chi$$

where for each $\chi \in \mathfrak{a}^*$, M_χ is the associated weight space

$$M_\chi = \{v \in M \mid \forall a \in \mathfrak{a},\, \exp(a) \cdot v = e^{\chi(a)} v\}.$$

The linear forms $\chi \in \mathfrak{a}^*$ for which $M_\chi \neq \{0\}$ are called the weights of M. Denote by $\Sigma(M)$ the sets of weights of M.

The set of of nontrivial weights of the adjoint representation of G is the set of restricted roots. We denote it by Σ. It forms a root system. We fix a set Σ_+ of positive roots and denote by $\mathfrak{a}^+$ the associated Weyl chamber:

$$\mathfrak{a}^+ = \{a \in \mathfrak{a} \mid \forall \alpha \in \Sigma_+,\, \alpha(a) \geq 0\}.$$

We also write $\mathfrak{a}^{++}$ to denote the interior of the Weyl chamber:

$$\mathfrak{a}^{++} = \{a \in \mathfrak{a} \mid \forall \alpha \in \Sigma_+,\, \alpha(a) > 0\}.$$

Let $g \in G$. The Cartan projection $\kappa(g)$ of g is the $\mathfrak{a}^+$-part in its Cartan decomposition, that is, the unique element in $\mathfrak{a}^+$ such that $g \in K \exp(\kappa(g))K$. The law of large numbers for a semisimple group, [6, Theorem 10.9], says that there is an element $\vec{\lambda}(\mu)$ in $\mathfrak{a}^{++}$, called the Lyapunov vector associated to μ, such that

$$\vec{\lambda}(\mu) = \lim_{n \to +\infty} \frac{1}{n} \int_G \kappa(g) \,\mathrm{d}\mu^{*n}(g).$$

If M is a simple G module, then M has a highest weight, denoted by $\chi_M \in \Sigma(M)$, so that for any weight $\chi \in \Sigma(M)$, $\chi_M - \chi$ is a sum of positive roots. By [6, Corollary 10.12], we have

$$\lambda_1(\mu, M) = \chi_M(\vec{\lambda}(\mu)).$$

Now let us recall the definition of the limit set of the group G in $\mathcal{M}_d(\mathbb{R})$. We write $\mathbb{R}G$ for the set of all elements $\mathcal{M}_d(\mathbb{R})$ of the form λg with $\lambda \in \mathbb{R}$, and $g \in G$. Let $\overline{\mathbb{R}G}$ denote the closure of $\mathbb{R}G$ in $\mathcal{M}_d(\mathbb{R})$ for the norm topology. Let r_G denote the proximal dimension of G, defined by

$$r_G = \min\bigl\{\operatorname{rank}(\pi) \mid \pi \in \overline{\mathbb{R}G} \setminus \{0\}\bigr\}. \tag{3.12}$$

The limit set of G in $\mathcal{M}_d(\mathbb{R})$ is defined to be

$$\Pi_G = \bigl\{\pi \in \overline{\mathbb{R}G} \mid \operatorname{rank}(\pi) = r_G\bigr\}.$$

Lemma 3.18. *Let $\mathbb{G} < \mathrm{SL}_d$ be connected semisimple $\mathbb{R}$-group. Assume that $G = \mathbb{G}(\mathbb{R})$ acts irreducibly on $\mathbb{R}^d$. Let $\pi_0 \in \mathcal{M}_d(\mathbb{R})$ be the spectral projector to the weight space associated to the highest weight. Then*

$$\Pi_G = \mathbb{R}^* K\pi_0 K = \mathbb{R}^* G\pi_0 G.$$

If moreover G is not compact then, writing $E = \mathrm{Span}_{\mathbb{R}}(G)$, the sum-set

$$\underbrace{G\pi_0 G + \cdots + G\pi_0 G}_{\dim E \text{ times}}$$

contains an open subset of E.

Proof. Let $\chi_0 = \chi_{\mathbb{R}^d} \in \Sigma(\mathbb{R}^d)$ denote the highest weight of the simple G-module $\mathbb{R}^d$. For a weight $\chi \in \Sigma(\mathbb{R}^d)$, let π_χ be the spectral projector to the associated weight space. Let $a \in \mathfrak{a}^{++}$ be any element. We have

$$\|\exp(na)\|^{-1}\exp(na) = \sum_{\chi \in \Sigma(\mathbb{R}^d)} e^{n(\chi(a)-\chi_0(a))}\pi_\chi.$$

Now by definition of the highest weight, $\chi(a) - \chi_0(a) < 0$ for $\chi \neq \chi_0$. It follows that

$$\pi_0 = \lim_{n\to+\infty} \|\exp(na)\|^{-1}\exp(na) \in \overline{\mathbb{R}G}.$$

Let $\pi \in \overline{\mathbb{R}G}$ be another nonzero element. There exists sequences $(\lambda_n) \in \mathbb{R}^{\mathbb{N}}$ and $(g_n) \in G^{\mathbb{N}}$ such that

$$\pi = \lim_{n\to+\infty} \lambda_n g_n.$$

Let $g_n = k_n \exp(a_n)\ell_n \in K\exp(\mathfrak{a}^+)K$ be the Cartan decomposition of g_n. By compactness of K, replacing (g_n) by a subsequence if necessary, we may assume that k_n converges to $k \in K$ and ℓ_n converges to ℓ. Then

$$k^{-1}\pi\ell^{-1} = \lim_{n\to+\infty} \lambda_n \exp(a_n)$$

Observe that

$$\exp(a_n) = \sum_{\chi\in\Sigma(\mathbb{R}^d)} e^{\chi(a_n)}\pi_\chi.$$

Hence

$$\lambda_n \exp(a_n) = \lambda_n e^{\chi_0(a_n)}(\pi_0 + \sum_{\chi\in\Sigma(\mathbb{R}^d)\setminus\{\chi_0\}} e^{\chi(a_n)-\chi_0(a_n)}\pi_\chi).$$

Note that $e^{\chi(a_n)-\chi_0(a_n)} \leq 1$ for all $\chi \in \Sigma(\mathbb{R}^d)$ and $n \geq 1$. We deduce that $\lambda_n e^{\chi_0(a_n)}$ converges to $\lambda \neq 0$, for otherwise π would be zero. Moreover,

$$\mathrm{rank}(\pi) = \mathrm{rank}(k^{-1}\pi\ell^{-1}) \geq \mathrm{rank}(\pi_0).$$

Equality holds if and only if $\lim_{n\to+\infty} e^{\chi(a_n)-\chi_0(a_n)} = 0$ for all $\chi \in \Sigma(\mathbb{R}^d)\setminus\{\chi_0\}$, which in turn is equivalent to

$$k^{-1}\pi\ell^{-1} = \lambda\pi_0.$$

Therefore, $r_G = \mathrm{rank}(\pi_0)$ and $\Pi_G \subset \mathbb{R}^*K\pi_0K$. We conclude by noticing that Π_G is invariant under multiplication by G on both sides : $\mathbb{R}^*K\pi_0K \subset \mathbb{R}^*G\pi_0G \subset \Pi_G$.

For the last assertion, assume that G is not compact. Then, $\chi_0 \neq 0$ and therefore $\chi_0(\mathfrak{a}) = \mathbb{R}$. For $a \in \mathfrak{a}$, $\exp(a)\pi_0 = e^{\chi_0(a)}\pi_0$, so that $\mathbb{R}_+^*\pi_0 \subset G\pi_0$ and hence

$$\mathbb{R}_+^* G\pi_0 G \subset G\pi_0 G.$$

Since the action of G on $\mathbb{R}^d$ is irreducible, E is an simple algebra over $\mathbb{R}$, by a version of Wedderburn's theorem [41, 2. page 194]. Observe that $\mathrm{Span}_{\mathbb{R}}(G\pi_0G)$ is a nontrivial two-sided ideal of E, hence $\mathrm{Span}_{\mathbb{R}}(G\pi_0G) = E$. Therefore we can pick

$\dim(E)$ elements $(\pi_1, \dots, \pi_{\dim(E)})$ from $G\pi_0 G$ making a basis of E. We conclude that

$$\underbrace{G\pi_0 G + \dots + G\pi_0 G}_{\dim E \text{ times}} \supset \mathbb{R}_+^* \pi_1 + \dots + \mathbb{R}_+^* \pi_{\dim(E)}$$

contains an open subset of E. $\square$

Proof of Lemma 3.17. The Lie algebra of G^k is $\mathfrak{g} \oplus \dots \oplus \mathfrak{g}$, in which we choose $\mathfrak{b} = \mathfrak{a} \oplus \dots \oplus \mathfrak{a}$ to be the Cartan subspace. Then the associated restricted root system is the direct sum $\Sigma \sqcup \dots \sqcup \Sigma \subset \mathfrak{b}^*$. We choose $\Sigma_+ \sqcup \dots \sqcup \Sigma_+$ as the set of positive roots so that $\mathfrak{b}^+ = \mathfrak{a}^+ \times \dots \times \mathfrak{a}^+$ is the corresponding Weyl chamber and

$$\vec{\lambda}(\mu^{\otimes k}) = (\vec{\lambda}(\mu), \dots, \vec{\lambda}(\mu)) \in \mathfrak{b}^+$$

is the Lyapunov vector associated to the random walk defined by $\mu^{\otimes k}$.

For any algebraic representation π of G^k, we denote by $\Sigma(G^k, \pi)$ the set of weights of π with respect to $\mathfrak{b}$.

Let $\sigma\colon G \to \mathrm{GL}(\mathbb{R}^d)$ denote the standard representation of G and, for $i = 1, \dots, k$, let $\sigma_i\colon G^k \to G \to \mathrm{GL}(\mathbb{R}^d)$ denote the representation of G^k obtained by composing the projection $G^k \to G$ to the i-th factor with σ. Note that for each i, there is a natural bijection between $\Sigma(G^k, \sigma_i) \to \Sigma(\sigma)$, $\chi \mapsto \tilde{\chi}$ such that the weight χ is the composition of the i-th projection with $\tilde{\chi} \in \Sigma(\sigma)$.

Let G^k act on $\mathbb{R}[\mathcal{M}_d^k]_{\leq D}$ by right translation. Let $\rho\colon G^k \to \mathrm{GL}(\mathbb{R}[\mathcal{M}_d^k]_{\leq D})$ denote the corresponding representation. Then, ρ is equivalent to

$$\bigoplus_{j=0}^{D} \mathrm{Sym}^j \big(\underbrace{\sigma_1 \oplus \dots \oplus \sigma_1}_{d \text{ times}} \oplus \dots \oplus \underbrace{\sigma_k \oplus \dots \oplus \sigma_k}_{d \text{ times}}\big).$$

It follows that any weight $\chi \in \Sigma(G^k, \rho)$ in ρ is the sum of at most D elements from $\bigcup_{i=1}^k \Sigma(G^k, \sigma_i)$. In particular,

$$\begin{aligned}
\lambda_1(\mu^{\otimes k}, \mathbb{R}[\mathcal{M}_d^k]_{\leq D}) &= \max_{\chi \in \Sigma(G^k, \rho)} \chi(\vec{\lambda}(\mu^{\otimes k})) \\
&\leq D \max_i \max_{\chi \in \Sigma(G^k, \sigma_i)} \chi(\vec{\lambda}(\mu^{\otimes k})) \\
&= D \max_i \max_{\chi \in \Sigma(G^k, \sigma_i)} \tilde{\chi}(\vec{\lambda}(\mu)) \\
&\leq D\lambda_1.
\end{aligned}$$

Since G is not compact, λ_1 is positive, by a result of Furstenberg [22], and it follows that

$$\lambda_1(\mu^{\otimes k}, \mathbb{R}[\mathcal{M}_d^k]_{< D}) < D\lambda_1.$$

The G^k-module $\mathbb{R}[\mathbb{G}^k]_{<D}$ is a quotient of $\mathbb{R}[\mathcal{M}_d^k]_{<D}$, whence

$$\lambda_1(\mu^{\otimes k}, \mathbb{R}[\mathbb{G}^k]_{<D}) < D\lambda_1. \tag{3.13}$$

Let $f_D \in \mathbb{R}[\mathcal{M}_d]$ denote the degree D homogeneous part of f. Then $\bar{F}_D \in \mathbb{R}[\mathcal{M}_d^k]$ defined by

$$\forall x_1, \dots, x_k \in \mathcal{M}_d,\ \bar{F}_D(x_1, \dots, x_k) = f_D(x_1 + \dots + x_k)$$

is the degree D homogeneous part of $\bar{F}$. Let $F_D \in$ be the restriction of $\bar{F}_D$ to $\mathbb{G}$. By (3.13) and the fact that $F - F_D \in \mathbb{R}[\mathbb{G}^k]_{<D}$, we get

$$p(F) = p(F_D).$$

We may therefore assume that f is homogeneous of degree D.

By Lemma 3.18, f does not vanish on $G\pi_0 G + \dots + G\pi_0 G$ where

$$\pi_0 = \lim_{n \to +\infty} \frac{\exp na}{\|\exp na\|} \in \mathcal{M}_d(\mathbb{R})$$

for some $a \in \mathfrak{a}^{++}$. Fix $g = (g_1, \dots, g_k) \in G^k$ and $h = (h_1, \dots, h_k) \in G^k$ such that

$$f(g_1\pi_0 h_1 + \dots g_k \pi_0 h_k) \neq 0.$$

Writing $b = (a, \dots, a) \in \mathfrak{b}$, we have, by the homogeneity of $\bar{F}$,

$$\lim_{n\to+\infty} \frac{F(g\exp(nb)h)}{\|\exp(na)\|^D} = \lim_{n\to+\infty} \bar{F}\Big(\frac{g\exp(nb)h}{\|\exp(na)\|}\Big) = f(g_1\pi_0 h_1 + \dots g_k\pi_0 h_k) \neq 0,$$

whence

$$|F(g\exp(nb)h)| \gg \|\exp(na)\|^D.$$

In the rest of the proof, the implied constants in the Vinogradov notation are independent of n but might depend on other quantities, like g or h. On the one hand,

$$\|\exp(na)\| \gg e^{n\chi_\sigma(a)},$$

and on the other hand,

$$|F(g\exp(nb)h)| = \big|\big(\rho(\exp(nb))\rho(h)F\big)(g))\big| \ll \|\rho(\exp(nb))\rho(h)F\|.$$

Decompose $\mathbb{R}[\mathbb{G}^k]_{\leq D} = \bigoplus_j M_j$ into simple G^k-modules and decompose $\rho(h)F = \sum_j F_j$ accordingly. Denote by χ_{M_j} the highest weight of M_j. Then

$$\|\rho(\exp(nb))\rho(h)F\| \leq \sum_j e^{n\chi_{M_j}(b)}\|F_j\|.$$

From the previous inequalities, there must exist j such that $F_j \neq 0$ and

$$e^{Dn\chi_\sigma(a)} \ll e^{n\chi_{M_j}(b)}\|F_j\|.$$

Hence $Dn\chi_\sigma(a) \leq n\chi_{M_j}(b) + O(1)$ and necessarily

$$D\chi_\sigma(a) \leq \chi_{M_j}(b).$$

We have seen that χ_{M_j} is the sum of D elements from $\bigcup_{i=1}^k \Sigma(G^k, \sigma_i)$:

$$\chi_{M_j} = \chi_1 + \dots + \chi_D.$$

Then

$$\chi_{M_j}(b) = \chi_1(b) + \dots + \chi_D(b) = \tilde{\chi}_1(a) + \dots + \tilde{\chi}_D(a).$$

Thus we have simultaneously

$$D\chi_\sigma(a) \leq \tilde{\chi}_1(a) + \dots + \tilde{\chi}_D(a)$$

and

$$\tilde{\chi}_1 + \dots + \tilde{\chi}_D \leq D\chi_\sigma$$

for the order over the set of weights. Since $a \in \mathfrak{a}^{++}$, , this forces

$$\tilde{\chi}_1 = \dots = \tilde{\chi}_D = \chi_\sigma.$$

Therefore

$$\lambda_1(\mu^{\otimes k}, M_j) = \chi_{M_j}(\vec{\lambda}(\mu^{\otimes k})) = (\tilde{\chi}_1 + \dots + \tilde{\chi}_D)(\vec{\lambda}(\mu)) = D\chi_\sigma(\vec{\lambda}(\mu)) = D\lambda_1.$$

We conclude that $p(\rho(h)F) \neq 0$ and hence

$$p(F) = \rho(h)^{-1}p(\rho(h)F) \neq 0. \qquad \square$$

We can now easily deduce Proposition 3.2 from Proposition 3.16 and Lemma 3.17.

Proof of Proposition 3.2. Note that under our assumptions, G cannot be compact. Let $f\colon E \to \mathbb{R}$ be a polynomial map of degree D, and denote by f_D its degree D homogeneous part. Define $F \in \mathbb{R}[\mathbb{G}^k]$ by

$$\forall g_1, \dots, g_k \in G,\ F(g_1, \dots, g_k) = f(g_1 + \dots + g_k).$$

Let $p\colon \mathbb{R}[\mathbb{G}^k]_{\leq D} \to \mathbb{R}[\mathbb{G}^k]_{\leq D}$ be the projection to the sum of all simple submodules M such that $\lambda_1(\mu^{\otimes k}, M) = D\lambda_1(\mu, \mathbb{R}^d)$. By Lemma 3.17, $\|p(F)\| = 0$ implies $\|f_D\| = 0$, and since these two expressions define seminorms on the space of polynomial maps on E, it follows that

$$\|f_D\| \ll \|p(F)\|. \tag{3.14}$$

By Proposition 3.16 applied to the random walk on G^k associated to the measure $\mu^{\otimes k}$, with $\lambda = D\lambda_1(\mu, \mathbb{R}^d)$ we get that for every $\omega > 0$, there exists $c > 0$ and $n_0 \in \mathbb{N}$ such that

$$\forall n \geq n_0, \quad \mu^{*n}\big(\{g \in G^k \mid |F(g)| \leq e^{(\lambda-\omega)n}\|p(F)\|\}\big) \leq e^{-cn}.$$

Together with (3.14), this proves what we want. □

3.8. **Fourier decay for random walks.** The relevant object here is the measure $\tilde{\mu}_n$, obtained from μ^{*n} after rescaling by a factor $e^{-\lambda_1 n}$. This rescaling shrinks μ^{*n} to a ball of subexponential size around 0. An important consequence of the results of this section and the previous one is the following theorem.

Theorem 3.19 (Fourier decay for $\tilde{\mu}_n$)**.** *Let μ be a probability measure on $\mathrm{SL}_d(\mathbb{Z})$, $d \geq 2$. Let Γ denote the subsemigroup generated by μ, $\mathbb{G}$ the Zariski closure of Γ in SL_d and E the subalgebra of $\mathcal{M}_d(\mathbb{R})$ generated by $\mathbb{G}(\mathbb{R})$. Denote $\tilde{\mu}_n = e^{-\lambda_1 n} * \mu^{*n}$ where λ_1 is the top Lyapunov exponent of μ. Assume that:*

- *(a) The measure μ has a finite exponential moment;*
- *(b) The action of Γ on $\mathbb{R}^d$ is irreducible;*
- *(c) The algebraic group $\mathbb{G}$ is Zariski connected.*

Then there exists $\alpha_0 > 0$ such that for every $0 < \alpha < \alpha_0$, there exists $c_0 > 0$ and $n_0 \geq 1$ such that for every $n \geq n_0$,

$$\forall \xi \in E^* \text{ with } e^{\alpha n} \leq \|\xi\| \leq e^{\alpha_0 n}, \quad |\widehat{\tilde{\mu}_n}(\xi)| \leq e^{-c_0 n}.$$

Proof. We want to apply Theorem 2.1 to the measure $\tilde{\mu}_n$, and for that, we should check that it is not concentrated near any affine subspace, nor near any translate of the set of non-invertible elements of E. This will follow from Propositions 3.1 and 3.2. Recall that given $\rho > 0$, we write $S_E(\rho)$ for the set

$$S_E(\rho) = \{x \in E \mid |\det_E(x)| \leq \rho\}.$$

Let $D = \dim(E)$. Under the assumptions of the theorem, we claim that there exists $\kappa > 0$ depending only μ such that for every $\omega > 0$, there exists $c = c(\mu, \omega) > 0$ such that for every $n \geq 1$, we can decompose the convolution

$$\tilde{\mu}_n^{\boxplus D} \boxminus \tilde{\mu}_n^{\boxplus D} = \eta + \theta$$

into positive Borel measures satisfying the following properties.

- (i) $\theta(E) \ll_\mu e^{-cn}$,
- (ii) $\eta(E \setminus B_E(0, e^{\omega n})) \ll_\mu e^{-cn}$,
- (iii) $\forall x \in E,\ \eta(x + S_E(e^{-\omega n})) \ll_\mu e^{-cn}$,
- (iv) $\forall \rho \geq e^{-n},\ \forall W < E$ affine subspace, $\eta(W^{(\rho)}) \ll_\mu e^{\omega n}\rho^\kappa$.

To justify this claim, let η_1 be the restriction of $\tilde{\mu}_n$ to $E \setminus B_E(0, e^{-\omega n})$ and put

$$\eta' = \eta_1 \boxplus \tilde{\mu}_n^{\boxplus(D-1)} \quad \text{and} \quad \eta = \eta' \boxminus \tilde{\mu}_n^{\boxplus D}.$$

By Theorem 3.10(i), there is $c = c(\mu, \omega) > 0$ such that for every $n \geq 1$,

$$\tilde{\mu}_n(B_E(0, e^{-\omega n})) \ll_\mu e^{-cn} \quad \text{and} \quad \tilde{\mu}_n(E \setminus B_E(0, e^{\omega n})) \ll_\mu e^{-cn}.$$

It follows that

$$\theta(E) \ll_\mu e^{-cn}$$

and

$$\eta\big(E \setminus B_E(0, 2De^{\omega n})\big) \ll_\mu e^{-cn}.$$

For $x \in E$, apply Proposition 3.2 to the polynomial function $y \mapsto \det_E(y - e^{\lambda_1 m} x)$. Note that these polynomials all have degree $D = \dim(E)$ and all have the same degree D homogeneous part, namely $\det_E$. We obtain $c > 0$ such that

$$\begin{aligned}\eta'(x + S_E(e^{-\omega n})) &\leq \mu_n^{\boxplus D}\big(\{y \in E \mid |\det_E(e^{-\lambda_1 n} y - x)| \leq e^{-\omega n}\}\big)\\ &\leq \mu_n^{\boxplus D}\big(\{y \in E \mid |\det_E(y - e^{\lambda_1 n} x)| \leq e^{(D\lambda_1 - \omega)m}\}\big)\\ &\ll_\mu e^{-cn}.\end{aligned}$$

Since this property is preserved under additive convolution, the same holds for η. Now let W be a proper affine subspace of E. Using the definition of η_1 and Proposition 3.1, we find for every $\rho \geq e^{-n}$,

$$\begin{aligned}\eta_1(\{g \in E \mid d(g, W) \leq \rho\}) &\leq \tilde{\mu}_n\big(\{g \in E \mid d(g, W) \leq \rho e^{\omega n}\|g\|\}\big)\\ &\leq \mu_n\big(\{g \in E \mid d(g, W) \leq \rho e^{\omega n}\|g\|\}\big)\\ &\ll_\mu e^{\omega\kappa n}\rho^\kappa.\end{aligned}$$

Again this property is preserved under additive convolution, so that η satisfies the required conditions.

Let $\varepsilon = \varepsilon(\mu, \kappa) > 0$ and $s = s(\mu, \kappa) \geq 1$ be the constant given by Theorem 2.1 applied with the parameter κ. Set $\omega = \alpha\varepsilon/2$. Recall that we chose $c = c(\mu, \omega)$. Finally set $\tau = \min\{c/2, \varepsilon\kappa\}$. With the choice of these parameters, for every n large enough, for every $R \in [e^{\alpha n}, e^n]$,

(i) $\theta(E) \leq R^{-\tau}$,
(ii) $\eta(E \setminus B_E(0, R^\varepsilon)) \leq R^{-\tau}$,
(iii) for all $x \in E$, $\eta(x + S_E(R^{-\varepsilon})) \leq R^{-\tau}$,
(iv) for all $\rho \geq R^{-1}$ and every proper affine subspace $W \subset E$, $\eta(W^{(\rho)}) \leq R^\varepsilon \rho^\kappa$.

In other words, the assumptions of Theorem 2.1 are satisfied for the measure η at the scale $\frac{1}{R}$. Therefore, for all $\xi \in E^*$ in the range

$$e^{\alpha n} \leq \|\xi\| \leq e^n,$$

we have

$$|\widehat{\eta^{*s}}(\xi)| \leq \|\xi\|^{-\varepsilon\tau}.$$

It follows that

$$|(\widehat{\tilde{\mu}_n^{\boxplus D} \boxminus \tilde{\mu}_n^{\boxplus D})^{*s}}(\xi)| \leq \|\xi\|^{-\varepsilon\tau} + O_s(\|\xi\|^{-\tau}).$$

Applying Lemma 2.7 to $\tilde{\mu}_n^{\boxplus D} \boxminus \tilde{\mu}_n^{\boxplus D}$, we get

$$|\widehat{\tilde{\mu}_{sn}}(\xi)| \leq \|\xi\|^{-c_0},$$

where $c_0 = \frac{\varepsilon\tau}{2(2D)^s}$ and, again, assuming that $n \geq n_0(\mu, \alpha)$. This shows the desired upper bound for $\widehat{\tilde{\mu}_n}(\xi)$ provided that n is a multiple of s.

To prove the estimate for general n, write $n = sq + r$ with $0 \leq r < s$. On the one hand, by (2.13),

$$\widehat{\tilde{\mu}_n}(\xi) = \int \widehat{\tilde{\mu}_{sq}}(\xi x)\, \mathrm{d}\tilde{\mu}_r(x),$$

On the other hand, it follows from the Markov inequality and the fact that $\tilde{\mu}_r$ has bounded exponential moment that, for x outside of a set of exponentially small $\tilde{\mu}_r$-measure,

$$e^{-\frac{\alpha n}{2}}\|\xi\| \leq \|\xi x\| \leq e^{\frac{n}{4s}}\|\xi\|.$$

This proves the theorem with $\alpha_0 = \frac{1}{4s}$. $\square$

4. The set of large Fourier coefficients

Starting from Theorem 3.19, we now use some Fourier analysis to show a first intermediate statement towards Theorem 1.2.

Proposition 4.1 (First step: concentration and separation)**.** *Let μ be a probability measure on $\mathrm{SL}_d(\mathbb{Z})$, $d \geq 2$. Denote by Γ the subsemigroup generated by μ, and by $\mathbb{G}$ the Zariski closure of Γ in SL_d. Assume that:*

(a) The measure μ has a finite exponential moment;
(b) The action of Γ on $\mathbb{R}^d$ is irreducible;
(c) The algebraic group $\mathbb{G}$ is Zariski connected.

There exist constants $C \geq 0$ and $\sigma > \tau > 0$ such that the following holds.
Let ν be a Borel probability measure on $\mathbb{T}^d$. Let $t_0 \in (0, 1/2)$. Assume that for some $a_0 \in \mathbb{Z}^d \setminus \{0\}$,

$$|\widehat{\mu^{*n} * \nu}(a_0)| \geq t_0 \quad \text{and} \quad n \geq C|\log t_0|.$$

Then, writing $N = e^{\sigma n}\|a_0\|$ and $M = e^{-\tau n}N$, there exists a $\frac{1}{M}$-separated set $X \subset \mathbb{T}^d$ such that

$$\nu\left(\bigcup_{x \in X} B(x, \frac{1}{N})\right) \geq t_0^C.$$

The proof of this statement goes by two steps. First, following [11], one applies a Fourier analytic lemma [11, Proposition 7.5] to translate the concentration of ν to a statement about its Fourier coefficients. Then, one uses the Fourier decay of $\tilde{\mu}_n$ to study the set of large Fourier coefficients

$$A(t) = \{a \in \mathbb{Z}^d \mid |\hat{\nu}(a)| \geq t\}, \tag{4.1}$$

and prove the desired statement.

4.1. **Detecting concentration from the Fourier coefficients.**

Because it is so elementary, and yet beautiful, we include the Fourier analytic lemma needed for our argument. The reader is referred to [11, Proposition 7.5] for its ingenious proof.

Lemma 4.2. *Given $d \in \mathbb{N}$, there exists $c > 0$ such that if a measure ν on $\mathbb{T}^d$ satisfies*

$$\mathcal{N}(A(t) \cap B(0, N), M) \geq s\left(\frac{N}{M}\right)^d$$

for some numbers $s, t > 0$ and some $M, N \geq 1$ such that $M < cN$, then there exists a $\frac{1}{M}$-separated subset $X \subset \mathbb{T}^d$ such that

$$\nu\left(\bigcup_{x \in X} B(x, \frac{1}{N})\right) \geq c(st)^3.$$

Going back to the statement of Proposition 4.1 above we see that it is enough to show that, under the same assumptions, there exist $C \geq 0$ and $\sigma > \tau > 0$ such that, for $N = e^{\sigma n}\|a_0\|$ and $M = e^{-\tau n}N$,

$$\mathcal{N}(A(t_0^C) \cap B(0, N), M) \geq t_0^C\left(\frac{N}{M}\right)^d. \tag{4.2}$$

This is the goal of the next paragraph.

4.2. **Fourier decay and large coefficients.** For $a \in \mathbb{Z}^d$ and $x \in \mathbb{T}^d$, we denote by $(a, x) \mapsto \langle a, x\rangle \in \mathbb{T}$ the natural pairing. Vectors in $\mathbb{Z}^d = \widehat{\mathbb{T}^d}$ indexing Fourier coefficients are naturally understood as row vectors, so that for any $g \in \mathrm{SL}_d(\mathbb{Z})$, we have

$$\langle a, gx\rangle = \langle ag, x\rangle.$$

Before we start the proof of (4.2), we record an elementary lemma – not much more than the Cauchy-Schwarz inequality – which shows that the set of large Fourier coefficients of a measure has some additive structure. It will later be combined with the multiplicative properties of μ^{*n}, allowing us to exploit the sum-product phenomenon for the study of the set of large Fourier coefficients. This approach to Fourier coefficients of multiplicative convolutions of measures goes back to the work of Bourgain and Konyagin [14] on exponential sums in finite fields.

We use the symbols $\boxplus$ and $\boxminus$ introduced in Section 2.

Lemma 4.3 (Additive structure of Fourier coefficients)**.** *Let μ be a Borel probability measure $\mathrm{SL}_d(\mathbb{Z})$ and ν a Borel probability measure on $\mathbb{T}^d$. If*

$$|\widehat{\mu * \nu}(a_0)| \geq t_0 > 0,$$

then for any integer $k \geq 1$, the set

$$A = \left\{g \in \mathcal{M}_d(\mathbb{Z}) \mid |\hat{\nu}(a_0 g)| \geq t_0^{2k}/2\right\}$$

satisfies

$$\left(\mu^{\boxplus k} \boxminus \mu^{\boxplus k}\right)(A) \geq \frac{t_0^{2k}}{2}.$$

Proof. Observe that

$$\widehat{\mu * \nu}(a_0) = \int_{\mathbb{T}^d}\int_\Gamma e(\langle a_0, gx\rangle)\,\mathrm{d}\mu(g)\,\mathrm{d}\nu(x) = \int_{\mathbb{T}^d}\int_\Gamma e(\langle a_0 g, x\rangle)\,\mathrm{d}\mu(g)\,\mathrm{d}\nu(x).$$

By Hölder's inequality,

$$\begin{aligned}
t_0^{2k} \leq |\widehat{\mu * \nu}(a_0)|^{2k} &\leq \int_{\mathbb{T}^d}\Big|\int_\Gamma e(\langle a_0 g, x\rangle)\,\mathrm{d}\mu(g)\Big|^{2k}\,\mathrm{d}\nu(x)\\
&\leq \int_{\Gamma^{2k}} \hat{\nu}\big(a_0(g_1 + \cdots + g_k - g_{k+1} - \cdots - g_{2k})\big)\,\mathrm{d}\mu^{\otimes 2k}(g_1, \ldots, g_{2k})\\
&\leq \int_E |\hat{\nu}(a_0 g)|\,\mathrm{d}\big(\mu^{\boxplus k} \boxminus \mu^{\boxplus k}\big)(g)\\
&\leq \big(\mu^{\boxplus k} \boxminus \mu^{\boxplus k}\big)(A) + \frac{t_0^{2k}}{2}\big(\mu^{\boxplus k} \boxminus \mu^{\boxplus k}\big)(E \setminus A),
\end{aligned}$$

which finishes the proof of the lemma. ∎

Combining the above observation and Theorem 3.19, we can derive (4.2).

Proof of (4.2). As before, for $n \geq 1$, we let $\tilde{\mu}_n = (e^{-\lambda_1 n})_*\mu_n$ denote the rescaling of $\mu_n = \mu^{*n}$. Denote by E the subalgebra of $\mathcal{M}_d(\mathbb{R})$ generated by $\mathbb{G}(\mathbb{R})$ and write $D = \dim(E)$. By Theorem 3.19 there exists constants $\alpha_0 > 0$ and $c_0 > 0$ such that

$$\forall \xi \in E^* \text{ with } e^{\frac{\alpha_0 n}{8D+4}} \leq \|\xi\| \leq e^{\alpha_0 n}, \quad |\widehat{\tilde{\mu}_n}(\xi)| \leq \|\xi\|^{-c_0},$$

Now fix $\delta = e^{-\frac{\alpha_0 n}{2}}$ and write $\alpha = 1/(4D+2)$. Let $C_1 = C_1(D, \alpha)$ from Lemma 4.4 below and set $k = \lceil C_1/c_0\rceil$ so that the above implies

$$\forall \xi \in E^* \text{ with } \delta^{-\alpha} \leq \|\xi\| \leq \delta^{-2}, \quad |\widehat{\tilde{\mu}_n^{\boxplus k} \boxminus \tilde{\mu}_n^{\boxplus k}}(\xi)| \leq \|\xi\|^{-C_1},$$

This says that the measure $\tilde{\mu}_n^{\boxplus k} \boxminus \tilde{\mu}_n^{\boxplus k}$ is regular at all scales between δ^2 and δ^α.

On the other hand, since $|\widehat{\mu_n * \nu}(a_0)| \geq t_0$, it follows from Lemma 4.3 that the set

$$A = \{g \in E \cap \mathcal{M}_d(\mathbb{Z}) \mid |\hat{\nu}(a_0 g)| \geq t_1 := t_0^{2k}/2\}$$

satisfies

$$\big(\mu_n^{\boxplus k} \boxminus \mu_n^{\boxplus k}\big)(A) \gg t_0^{2k}.$$

Letting $\tilde{A} = e^{-\lambda_1 n} \cdot A$ be the rescaling of A, we find

$$\big(\tilde{\mu}_n^{\boxplus k} \boxminus \tilde{\mu}_n^{\boxplus k}\big)(\tilde{A}) \gg t_0^{2k}.$$

From the large deviation estimate Theorem 3.10(i), we also have

$$\big(\tilde{\mu}_n^{\boxplus k} \boxminus \tilde{\mu}_n^{\boxplus k}\big)(E \setminus B_E(0, \delta^{-\alpha})) \ll_\mu e^{-c_2 n}$$

for some $c_2 = c_2(\mu) > 0$. Assuming $n \geq \frac{4k}{c_2}|\log t_0|$, this implies

$$\big(\tilde{\mu}_n^{\boxplus k} \boxminus \tilde{\mu}_n^{\boxplus k}\big)(\tilde{A} \cap B_E(0, \delta^{-\alpha})) \gg t_0^{2k}.$$

So we can apply Lemma 4.4 to the restriction of $\tilde{\mu}_n^{\boxplus k} \boxminus \tilde{\mu}_n^{\boxplus k}$ to $B_E(0, \delta^{-\alpha})$. Letting $t_1 = t_0^{2k}$, we obtain $x \in B_E(0, \delta^{-\alpha})$ such that

$$\mathcal{N}(\tilde{A} \cap B_E(x, \delta^{1/2}), \delta) \gg_D t_1^{D+1} \delta^{-D/2}.$$

Rescaling back, we find

$$\mathcal{N}\big(A \cap B_E(e^{\lambda_1 n} x, N_0), M_0\big) \gg_D t_2 \Big(\frac{N_0}{M_0}\Big)^D, \tag{4.3}$$

where $t_2 = t_1^{D+1}$, $N_0 = e^{\sigma n}$ and $M_0 = e^{-\tau n} N_0$ with $\sigma = \lambda_1 - \frac{\alpha_0}{4}$ and $\tau = \frac{\alpha_0}{4}$. In accordance with the statement of the proposition, we put

$$N = N_0 \|a_0\| \quad \text{and} \quad M = M_0 \|a_0\|.$$

Consider the map $\varphi_0 \colon E \to \mathbb{R}^d$, $g \mapsto a_0 g$. Letting $A' = A \cap B_E(e^{\lambda_1 n} x, N_0)$, we have

$$\mathcal{N}(A', M_0) \leq \mathcal{N}\big(\varphi_0(A'), M\big) \max_{b \in \varphi_0(A')} \mathcal{N}\big(A' \cap \varphi_0^{-1}(B(b, M)), M_0\big). \tag{4.4}$$

We claim that

$$\max_{b \in \varphi_0(A')} \mathcal{N}\big(A' \cap \varphi_0^{-1}(B(b, M)), M_0\big) \ll_E \Big(\frac{N_0}{M_0}\Big)^{D-d} \tag{4.5}$$

Evidently $\|\varphi_0\| \asymp_E \|a_0\|$. Let $W_0 = \ker \varphi_0$. Since G acts irreducibly on $\mathbb{R}^d$, φ_0 is surjective and hence $\dim(W_0) = D - d$. The restriction $\varphi_{0|W_0^\perp} \colon W_0^\perp \to \mathbb{R}^d$ is bijective. Moreover, by a compactness argument,

$$\|\varphi_{0|W_0^\perp}^{-1}\| \asymp_E \|a_0\|^{-1}.$$

Consequently, for any $y \in E$,

$$\varphi_0^{-1}(B(\varphi_0(y), M)) \subset y + W_0^{(O_E(M_0))}.$$

Hence

$$\mathcal{N}\big(A' \cap \varphi_0^{-1}(B(\varphi_0(y), M)), M_0\big) \leq \mathcal{N}\big(B_E(0, N_0) \cap W_0^{(O_E(M_0))}, M_0\big) \ll_E \Big(\frac{N_0}{M_0}\Big)^{D-d},$$

which proves the claim (4.5). From (4.3), (4.4) and (4.5), we get

$$\mathcal{N}\big(\varphi_0(A'), M\big) \gg_E t_2 \Big(\frac{N}{M}\Big)^d.$$

By definition of A', we have $\varphi_0(A') \subset A(t_1) \cap B(b, \|\varphi_0\| N_0)$, where $b = e^{\lambda_1 n} a_0 x$ and $\|\varphi_0\| N_0 \ll_E N$.

This is almost what we want, except that the ball $B(b, N)$ is not centered at the origin. To recenter that ball, we make use once more of the additive properties of the set of large Fourier coefficients. Choosing the densest ball of radius $N/2$ inside $B(b,N)$, we get some $b' \in \mathbb{R}^d$ such that

$$\mathcal{N}\big(A(t_1) \cap B(b', \frac{N}{2}), M\big) \gg_E t_2 \Big(\frac{N}{M}\Big)^d.$$

Choose an M-separated subset $A_1 \subset A(t_1) \cap B(b', N/2)$ of cardinality

$$|A_1| \gg \mathcal{N}\big(A(t_1) \cap B(b', \frac{N}{2}), M\big)$$

and such that all Fourier coefficients $\hat{\nu}(a)$, for $a \in A_1$, fall into the same quadrant of $\mathbb{C}$. Then

$$|A_1|\frac{t_1}{4} \leq \big|\sum_{a\in A_1} \hat{\nu}(a)\big| = \Big|\int_{\mathbb{T}^d} \sum_{a\in A_1} e(\langle a, x\rangle)\,\mathrm{d}\nu(x)\Big|$$

By the Cauchy-Schwarz inequality,

$$|A_1|^2\frac{t_1^2}{16} \leq \int_{\mathbb{T}^d} \big|\sum_{a\in A_1} e(\langle a, x\rangle)\big|^2\,\mathrm{d}\nu(x) = \sum_{a_1,a_2\in A_1} |\hat{\nu}(a_1 - a_2)|$$

Thus, there exists $a_2 \in A_1$ such that

$$|A_1|\frac{t_1^2}{16} \leq \sum_{a_1\in A_1} |\hat{\nu}(a_1 - a_2)|.$$

Set $A_2 = (A_1 - a_2) \cap A\big(\frac{t_1^2}{32}\big)$, we have $|A_2| \geq \frac{t_1^2}{32}|A_1|$ and $A_2 \subset B(0,N)$. It follows that

$$\mathcal{N}\Big(A\Big(\frac{t_1^2}{32}\Big) \cap B(0,N), M\Big) \gg_E t_1^2 t_2 \Big(\frac{N}{M}\Big)^d.$$

This concludes our proof. □

The next lemma is the regularity statement we need for measures on the Euclidean space that have a strong Fourier decay. It essentially states that if a set in $\mathbb{R}^D$ carries a large proportion of a measure with small Fourier coefficients at all frequencies between $\delta^{-\alpha}$ and $\delta^{-1-\alpha}$, then we can find a ball of radius $\delta^{O(\alpha)}$ in the set on which the measure is comparable to the Lebesgue measure at scale δ.

Lemma 4.4 (Regularity from Fourier decay)**.** *Given $D \geq 1$ and $\alpha > 0$, there exist constants $c = c(D,\alpha) > 0$ and $C_1 = C_1(D,\alpha) > 0$ such that the following holds for all $0 < \delta < ct$. Let μ be a Borel measure on $\mathbb{R}^D$, of total mass $\mu(\mathbb{R}^D) \leq 1$. Let A be a subset of $\mathbb{R}^D$. Assume*

(i) $\mathrm{Supp}(\mu) \subset B(0, \delta^{-\alpha})$,
(ii) $\forall \xi \in \mathbb{R}^D$*, with* $\delta^{-\alpha} \leq \|\xi\| \leq \delta^{-1-\alpha}$, $|\hat{\mu}(\xi)| \leq \|\xi\|^{-C_1}$,
(iii) $\mu(A) \geq t$.

Then there exists $x \in \mathbb{R}^D$ such that

$$\mathcal{N}(A \cap B(x, \delta^\beta), \delta) \geq ct^{D+1}\Big(\frac{\delta^\beta}{\delta}\Big)^D,$$

where $\beta = (2D+1)\alpha$.

Proof. Let $\varphi\colon \mathbb{R}^D \to \mathbb{R}$ be a nonnegative smooth function supported on $B(0,1)$ such that $\int_{\mathbb{R}^D} \varphi = 1$. Set $\varphi_\delta(x) = \delta^{-D}\varphi(\delta^{-1}x)$, $\forall x \in \mathbb{R}^D$. Note that

$$\forall \xi \in \mathbb{R}^D, \quad \widehat{\varphi_\delta}(\xi) = \hat{\varphi}(\delta\xi).$$

Since φ is smooth, for any $C_2 > 0$, we have

$$\forall \xi \in \mathbb{R}^D, \quad |\hat{\varphi}(\xi)| \ll_{C_2} (1 + \|\xi\|)^{-C_2}. \tag{4.6}$$

Define $\mu_\delta = \mu \boxplus \varphi_\delta$, viewed either as a measure or as a smooth function on $\mathbb{R}^D$. Clearly,

$$\mu_\delta(A^{(\delta)}) \geq \mu(A) \geq t.$$

Let $c > 0$ be a small constant depending on D and α to be determined later. Assume $\delta < ct$ and set $\rho = c\delta^\beta t$. Let $(B_i)_{1\leq i\leq i_{\max}}$ be an essentially disjoint covering of $B(0, \delta^{-\alpha})$ by closed balls of radius ρ. In other words, the intersection multiplicity of the covering is at most $C_d = O_d(1)$, so that in particular the number of balls is at most $i_{\max} = O_d(\delta^{D\alpha}\rho^{-D})$. Consider

$$I = \Big\{1 \leq i \leq i_{\max} \mid \frac{\mu_\delta(A^{(\delta)} \cap B_i)}{\mu_\delta(B_i)} \geq \frac{t}{2C_d}\Big\}.$$

Using the finite multiplicity of the covering, we infer that $\sum_{i\in I} \mu_\delta(B_i) \geq t/2$. Hence there exists $i \in I$ such that

$$\mu_\delta(B_i) \gg \frac{t}{i_{\max}} \gg_d t\delta^{D\alpha}\rho^{-D}.$$

We fix this i from now on. Define

$$M = \max_{x\in B_i} \mu_\delta(x) \quad \text{and} \quad m = \min_{x\in B_i} \mu_\delta(x).$$

We have $\mu_\delta(B_i) \leq M|B_i|$ and hence

$$M \gg_d t\delta^{D\alpha}. \tag{4.7}$$

Let $x_0 \in B_i$ such that $\mu_\delta(x_0) = M$. By the Plancherel theorem, for any $x \in B_i$,

$$\mu_\delta(x) = \mu \boxplus \varphi_\delta(x) = \int_{\mathbb{R}^D} e(-\langle\xi, x\rangle)\hat{\mu}(\xi)\widehat{\varphi_\delta}(\xi)\,\mathrm{d}\xi$$

Thus,

$$\begin{aligned} |\mu_\delta(x) - \mu_\delta(x_0)| &\leq \int_{\mathbb{R}^D} |1 - e(\langle\xi, x - x_0\rangle)||\hat{\mu}(\xi)||\widehat{\varphi_\delta}(\xi)|\,\mathrm{d}\xi \\ &\leq T_1 + 2T_2 + 2T_3, \end{aligned}$$

where

$$\begin{aligned} T_1 &= \int_{\|\xi\|\leq\delta^{-\alpha}} |1 - e(\langle\xi, x - x_0\rangle)|\,\mathrm{d}\xi \\ &\ll_d \int_{\|\xi\|\leq\delta^{-\alpha}} \|\xi\|\|x - x_0\|\,\mathrm{d}\xi \ll_d \delta^{-(D+1)\alpha}\rho, \\ T_2 &= \int_{\delta^{-\alpha}\leq\|\xi\|\leq\delta^{-1-\alpha}} |\hat{\mu}(\xi)|\,\mathrm{d}\xi \\ &\ll_d \int_{\delta^{-\alpha}\leq\|\xi\|\leq\delta^{-1-\alpha}} \|\xi\|^{-C_1}\,\mathrm{d}\xi \ll_d \frac{\delta^{(C_1-D)\alpha}}{C_1 - D}, \\ T_3 &= \int_{\|\xi\|\geq\delta^{-1-\alpha}} |\hat{\varphi}(\delta\xi)|\,\mathrm{d}\xi \\ &\ll_{D,C_2} \int_{\|\xi\|\geq\delta^{-1-\alpha}} \delta^{-C_2}\|\xi\|^{-C_2}\,\mathrm{d}\xi \ll_{D,C_2} \frac{\delta^{(C_2-D)\alpha-D}}{C_2 - D}. \end{aligned}$$

In the last line, we used (4.6).

Picking $\beta = (2D + 1)\alpha$, $C_1 = \frac{D\alpha+1}{\alpha} + D$ and $C_2 = \frac{D\alpha+D+1}{\alpha} + D$ and putting these inequalities together, we obtain, remembering (4.7) and $\delta < ct$,

$$M - m \ll_{D,\alpha} ct\delta^{D\alpha} + \delta^{D\alpha+1} \ll_{D,\alpha} cM.$$

This implies $M/m \leq 2$ provided that c is chosen small enough according to D and α. Remembering $i \in I$, we have

$$t \ll_d \frac{\mu_\delta(A^{(\delta)} \cap B_i)}{\mu_\delta(B_i)} \leq \frac{M}{m}\frac{|A^{(\delta)} \cap B_i|}{|B_i|}.$$

Hence

$$\mathcal{N}(A \cap B_i, \delta) \gg_d \delta^{-D} |A^{(\delta)} \cap B_i| \gg_d t\delta^{-D} |B_i| \gg_d t\delta^{-D} \rho^D = c^D t^{D+1} \delta^{D\beta - D}.$$

Let $x \in \mathbb{R}^D$ be the center of B_i. Then $A \cap B_i \subset A \cap B(x, \delta^\beta)$ and hence

$$\mathcal{N}(A \cap B(x, \delta^\beta), \delta) \gg_{D,\alpha} t^{D+1} \delta^{D\beta - D}. \qquad \square$$

5. Concentration near rational points

In this section, we finish the proof of Theorem 1.2 from the introduction. The argument follows closely the one given in Section 7 of [11], but some modifications are required since we cannot make use of the proximality assumption.

In all this section, unless stated otherwise, μ denotes a probability measure on $\mathrm{SL}_d(\mathbb{Z})$, $d \geq 2$. The Lyapunov exponents of μ are denoted by $\lambda_1 \geq \cdots \geq \lambda_d$. The subsemigroup generated by μ is denoted by Γ, its Zariski closure by $\mathbb{G}$, and we write $G = \mathbb{G}(\mathbb{R})$ for the set of real points. We assume that:

(a) The measure μ has a finite exponential moment;
(b) The action of Γ on $\mathbb{R}^d$ is irreducible;
(c) The algebraic group $\mathbb{G}$ is Zariski connected.

We also let E be the subalgebra of $\mathcal{M}_d(\mathbb{R})$ generated by G. For $n \in \mathbb{N}$, we write $\mu_n = \mu^{*n}$ for the law of the random walk in G at time n. Finally, ν denotes a Borel probability measure ν on $\mathbb{T}^d$, understood as the starting distribution of a random walk on $\mathbb{T}^d$. We write $\nu_n = \mu_n * \nu$ for the law of the random walk at time n.

We shall divide the proof of Theorem 1.2 into three parts. First, one observes that given a Borel probability measure ν on $\mathbb{T}^d$, the sequence of measures ν_n satisfies a diophantine property: if it gives much weight to a ball of small radius, then the ball must contain a rational point with small denominator. Second, starting the separated set X around which, by Proposition 4.1, ν_n is concentrated, one goes backwards along the random walk in order to increase the concentration of the measure around the set X, until one can apply the diophantine property to conclude that ν_{n-m} is concentrated near some rational points with bounded denominator. The last part, concluding the proof, is again going backwards along the random walk, to show that if ν_n concentrates near the set of rational points of bounded height, then ν is even more concentrated near that set.

5.1. **An almost diophantine property.**

The key to obtain the concentration near *rational* points is the following almost diophantine property of the sequence of measures $\nu_n = \mu_n * \nu$, $n \in \mathbb{N}$. Given $Q \geq 1$ and $\rho > 0$, we denote by W_Q the set of rational points in $\mathbb{T}^d$ with denominator at most Q, and by $\mathsf{W}_Q^{(\rho)}$ its ρ-neighborhood.

Proposition 5.1 (Almost diophantine property)**.** *Let μ be a probability measure on $\mathrm{SL}_d(\mathbb{Z})$, $d \geq 2$, with some finite exponential moment. Assume that μ acts strongly irreducibly on $\mathbb{R}^d$.*

There exist constants $C \geq 0$ and $\eta > 0$ depending only on μ, such that for every Borel probability measure ν on $\mathbb{T}^d$, for every $x \in \mathbb{T}^d$, every $\rho > 0$, and every $n \geq C|\log \rho|$,

$$\nu_n(B(x, \rho)) \geq \rho^\eta \quad \Longrightarrow \quad x \in \mathsf{W}_{\rho^{-1/10}}^{(\rho^{9/10})}.$$

Proof. By Lemma 3.14, the manifold $G \times G$ is not included in any proper affine subspace of $E \times E$ and therefore the set

$$\{g_1 + \cdots + g_d - h_1 - \cdots - h_d \mid g_i, h_i \in G\}$$

contains a non-empty open set in E. This implies in particular that the map $(g_i, h_i) \mapsto \det(g_1 + \cdots + g_d - h_1 - \cdots - h_d)$ is not identically zero on $G^d \times G^d$. By Proposition 3.8, we infer that there exists $c > 0$ such that for every m large enough,

$$\mu_m^{\otimes 2d}(\{(g_i, h_i)_{1 \le i \le d} \mid \det(\sum g_i - \sum h_i) = 0\}) \le e^{-cm}.$$

Set $\eta = \frac{c}{20d\lambda_1}$, $\rho \asymp e^{20d\lambda_1 m}$, and for $B = B(x, \rho)$, write

$$\begin{aligned}
e^{-cm} \asymp \rho^\eta &\le \nu_n(B)^{2d} \\
&= \Big(\int_{\mathbb{T}^d} \sum_{g\in\Gamma} \mu_m(g) \mathbb{1}_B(gx)\, \mathrm{d}\nu_{n-m}(x)\Big)^{2d} \\
&\le \int_{\mathbb{T}^d} \big(\sum_{g\in\Gamma} \mu_m(g) \mathbb{1}_{g^{-1}B}(x)\big)^{2d} \mathrm{d}\nu_{n-m}(x) \qquad \text{(by Jensen's inequality)} \\
&= \sum_{g_i, h_i} \mu_m(g_1)\cdots\mu_m(g_d)\mu_m(h_1)\cdots\mu_m(h_d)\nu_{n-m}(g_1^{-1}B \cap \cdots \cap h_d^{-1}B).
\end{aligned}$$

This shows that the $\mu_m^{\otimes 2d}$-measure of $2d$-tuples of elements $g_1, \dots, h_d$ such that

$$\nu_{n-m}(g_1^{-1}B \cap \cdots \cap h_d^{-1}B) \gg e^{-cm}$$

is at least e^{-cm}. In particular, using the large deviation estimate Theorem 3.10(i) and the observation above on the determinant, we may find elements $g_1, \dots, h_d$ in the support of μ_m satisfying

$$\begin{cases} \max(\|g_i\|, \|h_i\|) \le e^{2\lambda_1 m}, \\ \det(g_1 + \cdots + g_d - h_1 - \cdots - h_d) \ne 0, \\ g_1^{-1}B \cap \cdots \cap g_d^{-1}B \cap h_1^{-1}B \cap \cdots \cap h_d^{-1}B \ne \emptyset. \end{cases}$$

If $y \in \mathbb{R}^d$ represents a point in that intersection, then, writing $M = g_1 + \cdots + g_d - h_1 - \cdots - h_d$, there exists $v \in \mathbb{Z}^d$ such that

$$My = v + O(\rho),$$

whence

$$y = M^{-1}v + O(\|M\|^{-1}\rho). \tag{5.1}$$

Now, the matrix M has integer entries, and its determinant is bounded above by $e^{2d\lambda_1 m}$, so that the entries of M^{-1} are rational numbers with denominator bounded above by $e^{2d\lambda_1 m} \le \rho^{-\frac{1}{10}}$. Moreover,

$$\|M^{-1}\| \le \|M\|^{d-1} \det(M)^{-1} \le e^{2(d-1)\lambda_1 m}.$$

Equality (5.1) above shows that $x = g_1 y \mod \mathbb{Z}^d$ is at distance at most $\rho^{\frac{9}{10}}$ from a rational point with denominator at most $\rho^{-\frac{1}{10}}$. This finishes the proof. ☐

5.2. **Bootstrapping concentration.** We now wish to combine the diophantine property of $\nu_n = \mu_n * \nu$ with the concentration statement given by Proposition 4.1 to obtain some concentration near rational points. To help the reader follow our progress towards Proposition 5.7, we formulate another intermediate step, which is the goal of this paragraph.

Given a subset $X \subset \mathbb{T}^d$ and a small parameter $\rho > 0$, we shall write $X^{(\rho)}$ for the ρ-neighborhood of X. Thus, for instance, we write $\mathsf{W}_Q^{(\rho)}$ for the set of points in $\mathbb{T}^d$ that lie at distance at most ρ from a rational point with denominator at most Q.

Proposition 5.2 (Second step: concentration around rational points)**.** *Under the assumptions recalled at the beginning of this section, there exists a constant C depending only on μ such that the following holds.*
Let $t_0 \in (0, 1/2)$. Assume that for some $a_0 \in \mathbb{Z}^d \setminus \{0\}$,

$$|\widehat{\nu_n}(a_0)| \geq t_0 \quad \textit{and} \quad n \geq C \log \frac{\|a_0\|}{t_0}.$$

Then, for every integer m such that $m \geq C \log \frac{\|a_0\|}{t_0}$ and $n - m \geq Cm$,

$$\nu_{n-m}\big(\mathsf{W}_Q^{(Q^{-8})}\big) \geq t_0^C,$$

for some $Q \in [e^{\frac{m}{C}}, e^{Cm}]$.

The concentration statement given by Proposition 4.1 is not strong enough for a direct application of the diophantine property. We first need Lemma 5.3 below to bootstrap concentration. It is exactly the same statement as [11, Proposition 7.2], and the proof is also the same, with some minor modifications to avoid the use of the proximality assumption; we include it nonetheless, for readability.

Lemma 5.3. *Given $\varepsilon > 0$, there exist $c > 0$ and $m_0 \in \mathbb{N}$ so that for $m \geq m_0$, the following holds for every Borel probability measure ν on $\mathbb{T}^d$. Given scales $r, \rho > 0$ such that $e^{d\lambda_1 m}\rho < r$, there are scales $r_1 = e^{-(\lambda_1+\varepsilon)m} r$ and $\rho_1 = e^{-(\lambda_1-\varepsilon)m}\rho$, so that for every r-separated set $X \subset \mathbb{T}^d$, one can construct an r_1-separated set $X_1 \subset \mathbb{T}^d$ with*

$$\nu(X_1^{(\rho_1)}) \geq \nu_m(X^{(\rho)})^d - e^{-cm}.$$

Proof. First, by Jensen's inequality used in the same way as in the proof of Proposition 5.1,

$$\nu_m(X^{(\rho)})^d \leq \sum_{g_1,\dots,g_d \in \Gamma} \mu_m(g_1)\dots\mu_m(g_d)\nu(g_1^{-1}X^{(\rho)} \cap \dots \cap g_d^{-1}X^{(\rho)}).$$

This implies that the set of d-tuples $(g_i)_{1\leq i\leq d}$ such that

$$\nu(g_1^{-1}X^{(\rho)} \cap \dots \cap g_d^{-1}X^{(\rho)}) \geq \nu_m(X^{(\rho)})^d - e^{-cm} \tag{5.2}$$

has $\mu_m^{\otimes d}$-measure at least e^{-cm}. By Theorem 3.10(ii) and Lemma 5.4 below, if c is chosen small enough, there must exist $(g_1, \dots, g_d)$ satisfying this inequality, and moreover

$$\forall i = 1, \dots, d, \quad \|g_i\| \leq e^{(\lambda_1+\varepsilon)m} \quad \text{and} \quad \|g_i^{-1}\| \leq e^{(-\lambda_d+\varepsilon)m} \tag{5.3}$$

and

$$\forall v \in \mathbb{R}^d \setminus \{0\}, \quad \max_i \frac{\|g_i v\|}{\|v\|} \geq e^{(\lambda_1-\varepsilon)m}. \tag{5.4}$$

We fix such elements $g_1, \dots, g_d$ for the rest of the proof.

Without loss of generality, we may assume that $\varepsilon > 0$ is so small that

$$\lambda_1 - \lambda_d + 3\varepsilon < d\lambda_1.$$

We claim then that the set $g_1^{-1}X^{(\rho)} \cap \dots \cap g_d^{-1}X^{(\rho)}$ is included in a union of at most $|X|$ balls of radius $\rho_1 = e^{-(\lambda_1-\varepsilon)m}\rho$: Indeed, from (5.3) one finds – drawing a picture of $X^{(\rho)}$ and $g_i^{-1}X^{(\rho)}$ – that given $x \in X$ and $i \geq 1$, the set $g_1^{-1}B(x, \rho)$ meets at most one component $g_i^{-1}B(y, \rho)$, $y \in X$. Therefore, there are at most $|X|$ non-empty intersections $g_1^{-1}B(x_1, \rho) \cap \dots g_d^{-1}B(x_d, \rho)$, for $x_1, \dots, x_d \in X$.

If x, y lie inside such an intersection, then, for each i, $\|g_i(x - y)\| \leq \rho$, and (5.4) implies that $\|x-y\| \leq e^{-(\lambda_1-\varepsilon)m}\rho = \rho_1$. Thus, each intersection $g_1^{-1}B(x_1, \rho) \cap \dots \cap g_d^{-1}B(x_d, \rho)$ is included in a ball of radius ρ_1.

Finally, using (5.3) again, we see that these intersections are separeated by at least $r_1 = e^{-(\lambda_1+\varepsilon)m}r$, and the proposition follows. □

We now state and prove the large deviation estimate use in the above argument.

Lemma 5.4. *Let μ be a Borel probability measure on $\mathrm{SL}_d(\mathbb{R})$ with some finite exponential moment, and assume that the semigroup Γ generated by μ acts strongly irreducibly on $\mathbb{R}^d$. Then, for every $\varepsilon > 0$, there exists $c > 0$ such that for every large enough $m \in \mathbb{N}$,*

$$\mu_m^{\otimes d}\Big\{(g_1,\dots,g_d) \mid \forall v \in \mathbb{R}^d \setminus \{0\}, \max_i \frac{\|g_i v\|}{\|v\|} \geq e^{(\lambda_1-\varepsilon)m}\Big\} \geq 1 - e^{-cm}.$$

Remark 4. If one assumes that μ is supported on $\mathrm{SL}_d(\mathbb{Z})$, and replaces d by d^2, then this lemma follows directly from Proposition 3.2. This particular case would be sufficient for our purposes.

Proof. In this proof, c denotes a small positive constant, depending on ε, and whose value may vary from one line to the other. Let r denote the proximality dimension of Γ. We shall use the notation introduced in the proof of Proposition 3.11. Recall that for $g \in \Gamma$, we consider its Cartan decomposition $g = k\,\mathrm{diag}(\sigma_1(g),\dots,\sigma_d(g))\ell$, where k and ℓ are orthogonal matrices and $\sigma_1(g) \geq \dots \geq \sigma_d(g)$ are the singular values of g. We defined

$$W_g^- = \ell^{-1}\,\mathrm{Span}(e_{r+1},\dots,e_d)$$

where $(e_1,\dots,e_d)$ is the standard basis of $\mathbb{R}^d$, so that for every non-zero $v \in \mathbb{R}^d$,

$$\frac{\|gv\|}{\|v\|} \geq \mathrm{d}_\measuredangle(\mathbb{R}v, W_g^-)\|g\|, \tag{5.5}$$

By the large deviation estimate Theorem 3.10(i) if $g_1,\dots,g_d$ are independent random variables with law μ_m, then with probability at least $1-e^{-cm}$,

$$\forall i \in \{1,\dots,d\}, \quad \|g_i\| \geq e^{(\lambda_1-\varepsilon)m}.$$

For a subspace $W \leq \mathbb{R}^d$, we let $\mathrm{Nbd}(W,\rho)$ denote the ρ-neighborhood of W in $\mathbb{R}^d$. It follows from the above that the lemma will be proved – with $4d\varepsilon$ instead of ε – if we can show that with probability at least $1-e^{-cm}$, the intersection

$$\bigcap_{i=1}^d \mathrm{Nbd}(W_{g_i}^-, e^{-4d\varepsilon m})$$

reduces to a ball of radius $\frac{1}{2}$.

For that, we construct inductively for $k = 1,\dots,d-r+1$ a linear subspace W_k of dimension $d-r+1-k$, depending on $g_1,\dots,g_k$, such that

$$\bigcap_{i=1}^k \mathrm{Nbd}(W_{g_i}^-, e^{-4d\varepsilon m}) \subset \mathrm{Nbd}(W_k, e^{-4(d+1-k)\varepsilon m}).$$

At each step W_{k+1} is constructed in terms of W_k and g_{k+1} and the construction is possible with probability $1-e^{-cm}$. For $k=1$, one may simply take $W_1 = W_{g_1}^-$. Then, suppose W_k has been constructed, and let $\mathbb{R}w \subset W_k$ be any line. By Theorem 3.10(iii), with probability at least $1-e^{-cm}$

$$\frac{\|g_{k+1}w\|}{\|w\|} \geq e^{(\lambda_1-\varepsilon)m}.$$

By Theorem 3.10(ii), with probability $1-e^{-cm}$,

$$\forall j \in \{1,\dots,d\}, \quad |\frac{1}{m}\log\sigma_j(g_{k+1}) - \lambda_j| \leq \varepsilon, \tag{5.6}$$

and by a straightforward generalization of [11, Lemma 4.1(2)],

$$\frac{\|gw\|}{\|w\|} \leq \|g\|d(\mathbb{R}w, W_{g_{k+1}}^-) + \sigma_{r+1}(g_{k+1}).$$

Since by a theorem of Guivarc'h and Raugi [23], $\lambda_r > \lambda_{r+1}$, we deduce from the above that, provided $\varepsilon > 0$ is small enough,

$$d(\mathbb{R}w, W^-_{g_{k+1}}) \geq e^{-3\varepsilon m}.$$

This implies that there exists a proper subspace $W_{k+1} < W_k$ such that

$$\mathrm{Nbd}(W_k, e^{-4(d-k+1)\varepsilon m}) \cap \mathrm{Nbd}(W^-_{g_{k+1}}, e^{-4d\varepsilon m}) \subset \mathrm{Nbd}(W_{k+1}, e^{-4(d-k)\varepsilon m}).$$

This proves what we want. □

To prove Proposition 5.2, we proceed as follows. Applying first Proposition 4.1, we shall obtain $m_0 \in \mathbb{N}$ and scales ρ_0 and r_0, together with an r_0-separated set $X_0 \subset \mathbb{T}^d$ such that

$$\nu_{n-m_0}(X_0^{(\rho_0)}) \geq t_0^C.$$

The idea is then to reduce the radius of the balls, by an iterated application of Lemma 5.3. We shall thus obtain an increasing sequence of integers m_k and a decreasing sequence of scales ρ_k and r_k, $k = 1, 2, \dots$ together with an r_k-separated set $X_k \subset \mathbb{T}^d$ such that $|X_k| \leq |X_0|$ and

$$\nu_{n-m_k}(X_k^{(\rho_k)}) \geq t_0^{C_k}.$$

Once we arrive at a scale ρ_k such that

$$|X_k| \leq |X_0| \leq \rho_k^{-\frac{\eta}{2}},$$

we shall be able to use the diophantine property of the random walk to conclude. Now let us turn to the detailed proof.

Proof of Proposition 5.2. Let C_0 and $\sigma > \tau > 0$ be the constants given by Proposition 4.1. and C' and $\eta > 0$ the ones given by Proposition 5.1. Then write

$$m = m_0 + km_+,$$

where

$$m_0 \geq C_0 |\log t_0|, \qquad \frac{\tau m_0}{4d\lambda_1} \leq m_+ \leq \frac{\tau m_0}{2d\lambda_1}$$

and

$$k = \lceil \frac{16 d^2 \sigma}{\eta \tau} \rceil = O_\mu(1).$$

This is feasible provided $m \geq C|\log t_0|$, where $C \geq 0$ depends on μ via the constants C_0, τ, etc. Note that within constants depending only on μ,

$$m \asymp m_0 \asymp m_+.$$

By Proposition 4.1 applied to

$$\nu_n = \mu_{m_0} * \nu_{n-m_0},$$

there exist scales

$$\rho_0 = e^{-\sigma m_0} \|a_0\|^{-1} \quad \text{and} \quad r_0 = e^{\tau m_0} \rho_0$$

together with an r_0-separated subset $X_0 \subset \mathbb{T}^d$ such that

$$\nu_{n-m_0}(X_0^{(\rho_0)}) \geq t_0^{C_0}.$$

Note that, since X_0 is r_0-separated,

$$|X_0| \ll_d r_0^{-d} \leq e^{d(\sigma-\tau)m_0} \|a_0\|^d.$$

Thus if C was chosen large enough, we have

$$|X_0| \leq e^{d\sigma m_0}.$$

Choose $\varepsilon > 0$ such that $2k\varepsilon < d\lambda_1$ so that

$$k\varepsilon m_+ < \frac{\tau m_0}{4}$$

and apply Lemma 5.3 to

$$\nu_{n-m_0} = \mu_{m_+} * \nu_{n-m_0-m_+}.$$

This is allowed since by our choice of parameters

$$e^{d\lambda_1 m_+}\rho_0 \leq e^{\frac{\tau m_0}{2}}\rho_0 < r_0.$$

This yields scales

$$\rho_1 = e^{-(\lambda_1-\varepsilon)m_+}\rho_0 \quad \text{and} \quad r_1 = e^{-(\lambda_1+\varepsilon)m_+}r_0$$

together with an r_1-separated subset X_1 such that $|X_1| \leq |X_0|$ and

$$\nu_{n-m_0-m_+}(X_1^{(\rho_1)}) \geq t_0^{C_1},$$

provided m is large enough to ensure that $e^{-cm_+} < t_0^{C_1}$. We may repeat this procedure at least k times, and therefore obtain a sequence of scales defined inductively by

$$\rho_{i+1} = e^{-(\lambda_1-\varepsilon)m_+}\rho_i \quad \text{and} \quad r_{i+1} = e^{-(\lambda_1+\varepsilon)m_+}r_i.$$

Indeed, our choice of ε ensures that for every $i \leq k$,

$$e^{d\lambda_1 m_+}\rho_i \leq e^{d\lambda_1 m_+ + 2i\varepsilon m_+ - \tau m_0} r_i < r_i.$$

In the end, we obtain scales ρ_k and r_k, and a set X_k with

$$|X_k| \leq |X_0| \leq e^{d\sigma m_0}$$

such that

$$\nu_{n-m}(X_k^{(\rho_k)}) \geq t_0^{C_k}. \tag{5.7}$$

Moreover,

$$\rho_k = e^{-k(\lambda_1-\varepsilon)m_+}\rho_0 \leq e^{-\frac{k\lambda_1 m_+}{2}} \leq e^{-\frac{2d\sigma m_0}{\eta}}$$

so that

$$|X_k| \leq \rho_k^{-\frac{\eta}{2}}.$$

Therefore, adjusting slightly the values of the constants, we may restrict X_k to the points satisfying

$$\nu_{n-m}(B(x,\rho_k)) \geq \rho_k^{\eta},$$

while preserving (5.7).

Note that we also have $\rho_k^{-1} \leq e^{(\sigma+\lambda_1)m}\|a_0\|$. Thus, if C was chosen large enough, then $n - m \geq Cm \geq C'|\log\rho_k|$, and we may conclude by Proposition 5.1 that

$$X_k^{(\rho_k)} \subset \mathsf{W}_{\rho_k^{-1/10}}^{(\rho_k^{8/10})}.$$

This proves the proposition with $Q = \rho_k^{-1/10}$ in the desired range. □

5.3. **End of the proof of Proposition 5.7: near rational points.** The end of the proof of Proposition 5.7 is based on an argument similar in spirit to the one used in Lemma 5.3, to bootstrap concentration. The proposition we shall need is again taken from [11], where it appears as [11, Proposition 7.4]. The proof we present follows closely the one given in [11], but the key Lemma 5.6 below, analogous to [11, Lemma 7.10], is proved using a new argument, which avoids using a regularity property of the μ-stationary measure on the projective space, only available with a proximality assumption.

Proposition 5.5. *Let μ be a probability measure on $\mathrm{SL}_d(\mathbb{Z})$ with some finite exponential moment and acting strongly irreducibly on $\mathbb{R}^d$. Given $\varepsilon > 0$, there exist m_* and $\omega > 0$ such that if $\rho > 0$, $Q \geq 1$ and $m \geq m_*$ satisfy*

$$e^{d\lambda_1 m}\rho < Q^{-2},$$

and ν is any Borel probability measure on $\mathbb{T}^d$, then

$$(\mu_m * \nu)(\mathsf{W}_Q^{(e^{-(\lambda_1-\varepsilon)m}\rho)}) \geq \nu(\mathsf{W}_Q^{(\rho)}) - e^{-\omega m}.$$

The proof of this proposition is based on the following lemma.

Lemma 5.6. *Let μ be a Borel probability measure on $\mathrm{SL}_d(\mathbb{R})$ with some finite exponential moment and whose support generates a subsemigroup acting strongly irreducibly on $\mathbb{R}^d$. Given $\varepsilon > 0$, there exists $\theta > 0$ such that the following holds for every integer m sufficiently large.*

Let A be a subset of $\mathrm{SL}_d(\mathbb{R})$ such that $\mu_m(A) \geq e^{-\theta m}$. There exists a subset $\mathcal{G} = \{g_i\}_{1\leq i\leq k}$ of cardinality $k \geq e^{\theta m}$ in A such that for every subset $\{g_{i_1}, \dots, g_{i_d}\}$ of d elements of $\mathcal{G}$, for every v in $\mathbb{R}^d$,

$$\max_{1\leq j\leq d} \|g_{i_j} v\| \geq e^{(\lambda_1-\varepsilon)m}\|v\|.$$

Proof. Having fixed $\varepsilon > 0$, let

$$\mathcal{T} = \{(h_1, \dots, h_d) \mid \forall v \in \mathbb{R}^d \setminus \{0\}, \max_i \frac{\|h_i v\|}{\|v\|} < e^{(\lambda_1-\varepsilon)m}\}.$$

By Lemma 5.4, there exists $c > 0$ such that for every large enough m,

$$\mu_m^{\otimes d}(\mathcal{T}) \leq e^{-cm}.$$

We shall prove that the lemma holds with $\theta = \frac{c}{d2^{d+1}}$. Let

$$A_1 = \{g \in A \mid \mu_m^{\otimes d-1}(\{(h_2, \dots, h_d) \mid (g, h_2, \dots, h_d) \in \mathcal{T}\}) \geq e^{-cm/2}\}.$$

Then

$$e^{-cm} \geq \mu_m^{\otimes d}(\mathcal{T}) \geq e^{-cm/2}\mu_m(A_1),$$

and therefore

$$\mu_m(A_1) \leq e^{-cm/2}.$$

To construct $\mathcal{G}$, we first choose $g_1 \in A \setminus A_1$; this is possible because $\theta < c/2$. Let

$$A_2(g_1) = \{g \in A \mid \mu_m^{\otimes d-2}(\{(h_3, \dots, h_d) \mid (g_1, g, h_3 \dots, h_d) \in \mathcal{T}\}) \geq e^{-cm/4}\}.$$

Since $g_1 \notin A_1$, we have

$$\begin{aligned} e^{-cm/2} &\geq \mu_m^{\otimes d-1}(\{(h_2, \dots, h_d) \mid (g_1, h_2, \dots, h_d) \in \mathcal{T}\}) \\ &\geq e^{-cm/4}\mu_m(A_2(g_1)), \end{aligned}$$

whence

$$\mu_m(A_2(g_1)) \leq e^{-cm/4}.$$

We may therefore pick an element $g_2 \in A$ such that $g_2 \notin A_1 \cup A_2(g_1)$. Then set

$$A_3(g_1, g_2) = \{g \mid \mu_m^{\otimes d-3}(\{(h_4, \dots, h_d) \mid (g_1, g_2, g, h_4, \dots, h_d) \in \mathcal{T}\}) \geq e^{-cm/8}\}$$

for which it is readily checked, using the fact $g_2 \notin A_1(g_1)$, that

$$\mu_m(A_3(g_1, g_2)) \leq e^{-cm/8}.$$

This allows us to pick $g_3 \in A$ such that $g_3 \notin A_1 \cup A_2(g_1) \cup A_2(g_2) \cup A_3(g_1, g_2)$. Following this procedure, the elements $g_1, g_2, g_3, \dots$ of $\mathcal{G}$ are constructed inductively. Once $g_1, \dots, g_k$ have been chosen, one picks $g_{k+1} \in A$ outside the union of all subsets

$$A_r(g_{i_1}, \dots, g_{i_{r-1}}) = \{g \mid \mu_m^{\otimes d-r}(\{(h_{r+1}, \dots, h_d) \mid (g_{i_1}, \dots, g_{i_{r-1}}, g, h_{r+1}, \dots, g_d) \in \mathcal{T}\}) \geq e^{-cm/2^r}\},$$

where $(g_{i_1}, \dots, g_{i_{r-1}})$ can be any subset of $(g_1, \dots, g_k)$ with at most d elements. By convention, for $r = d$, write

$$A_d(g_{i_1}, \dots, g_{i_{d-1}}) = \{g \mid (g_{i_1}, \dots, g_{i_{d-1}}, g) \in \mathcal{T}\}.$$

Just as above, one checks by induction, using $g_{i_{r-1}} \notin A_{r-1}(g_{i_1}, \dots, g_{i_{r-2}})$, that

$$\mu_m(A_r(g_{i_1}, \dots, g_{i_{r-1}})) \leq e^{-cm/2^r} \leq e^{-cm/2^d}.$$

Thus, at step k, the union of all subsets $A_r(g_{i_1}, \dots, g_{i_r})$ to be avoided has measure at most

$$[1 + \binom{k}{1} + \dots + \binom{k}{d}]e^{-cm/2^d} \ll k^d e^{-cm/2^d}.$$

So the procedure can go on as long as $k^d e^{-cm/2^d} < \mu_m(A)$. Since $\mu_m(A) \geq e^{-\theta m} = e^{-cm/2^{d+1}}$, one can at least reach some $k \geq e^{\frac{cm}{d2^{d+1}}}$, which proves the lemma. $\square$

The rest of the proof of Proposition 5.5 is exactly as in [11, §7.D.]; we include it for completeness.

Proof of Proposition 5.5. Once more, write

$$\nu_m(\mathsf{W}_Q^{(\rho)}) = \sum_g \mu_m(g)\nu(g^{-1}\mathsf{W}_Q^{(\rho)}),$$

to observe that

$$A = \{g \mid \nu(g^{-1}\mathsf{W}_Q^{(\rho)}) \geq (\mu_m * \nu)(B) - e^{-\theta m}\}$$

satisfies

$$\mu_m(A) \geq e^{-\theta m}.$$

Using the large deviation estimate for $\|g^{-1}\|$, we may reduce A without any significant loss of μ_m-measure so that for every g in A,

$$\|g^{-1}\| \leq e^{(\lambda_d + \varepsilon)m} \leq \frac{1}{2} e^{d\lambda_1 m}.$$

By Lemma 5.6, there exists a subset $\mathcal{G} \subset A$ of cardinality at least $e^{\theta m}$ such that for any distinct elements $g_1, \dots, g_d$ in $\mathcal{G}$, for every $v \in \mathbb{R}^d$,

$$\max_{1 \leq i \leq d} \|g_i v\| \geq e^{(\lambda_1 - \varepsilon)m} \|v\|.$$

For such elements $g_1, \dots, g_d$,

$$g_1^{-1}\mathsf{W}_Q^{(\rho)} \cap \dots g_d^{-1}\mathsf{W}_Q^{(\rho)} = \bigcup_{x_1, \dots, x_d \in \mathsf{W}_Q} g_1^{-1}B(x_1, \rho) \cap \dots \cap g_d^{-1}B(x_d, \rho).$$

Now, if $u \in \mathbb{R}^d$ represents an element of $g_1^{-1}B(x_1, \rho) \cap \dots \cap g_d^{-1}B(x_d, \rho)$, then, for some vectors $v_i \in \mathbb{Z}^d$,

$$g_i u = x_i + v_i + O(\rho), \quad i = 1, \dots, d \tag{5.8}$$

i.e.

$$u = g_i^{-1}(x_i + v_i) + O(e^{d\lambda_1 m}\rho).$$

But the points $g_i^{-1}(x_i + v_i)$ are rational with denominator at most Q, so that they are at least Q^{-2} away from one another. Since $e^{d\lambda_1 m}\rho < Q^{-2}$, this shows that there exists $u_0 \in \mathbb{R}^d$, rational with denominator at most Q such that for each i, $g_i u_0 = x_i + v_i$. Coming back to (5.8) above, we find

$$\|g_i(u - u_0)\| \leq \rho, \quad i = 1, \dots, d$$

and by definition of the subset $\mathcal{G}$,

$$\|u - u_0\| \leq e^{-m(\lambda_1 - \varepsilon)}\rho.$$

This shows that
$$g_1^{-1}\mathsf{W}_Q^{(\rho)} \cap \dots g_d^{-1}\mathsf{W}_Q^{(\rho)} \subset \mathsf{W}_Q^{(e^{-(\lambda_1-\varepsilon)m}\rho)}.$$
In other words, the family of subsets
$$\{g^{-1}\mathsf{W}_Q^{(\rho)} \setminus \mathsf{W}_Q^{(e^{-m(\lambda_1-\varepsilon)}\rho)}\ ;\quad g\in\mathcal{G}\}$$
has intersection multiplicity less than d. Therefore,
$$\sum_{g\in\mathcal{G}} \nu_m(g^{-1}\mathsf{W}_Q^{(\rho)} \setminus \mathsf{W}_Q^{(e^{-m(\lambda_1-\varepsilon)}\rho)}) \leq d,$$
and as $|\mathcal{G}| \geq e^{\theta m}$, there must exist g in $\mathcal{G}$ such that
$$\nu_m(g^{-1}\mathsf{W}_Q^{(\rho)} \setminus\ \mathsf{W}_Q^{(e^{-m(\lambda_1-\varepsilon)}\rho)}) \leq de^{-\theta m}.$$
Then,
$$\begin{aligned}\nu_m(\mathsf{W}_Q^{(e^{-m(\lambda_1-\varepsilon)}\rho)}) &\geq \nu_m(g^{-1}\mathsf{W}_Q^{(\rho)}) - de^{-\theta m}\\ &\geq \nu(\mathsf{W}_Q^{(\rho)}) - (d+1)e^{-\theta m}\\ &\geq \nu(\mathsf{W}_Q^{(\rho)}) - e^{-\omega m}.\end{aligned}$$
□

We are finally ready to prove the main theorem of this article, Theorem 1.2, announced in the introduction. We shall in fact prove a slightly more general statement, given as Proposition 5.7 below. Recall that for parameters $Q \geq 1$ and $\rho > 0$, we write W_Q for the set of rational points on $\mathbb{T}^d$ with denominator at most Q, and $\mathsf{W}_Q^{(\rho)}$ for its ρ-neighborhood.

Proposition 5.7. *Let $d \geq 2$. Let μ be a probability measure on $\mathrm{SL}_d(\mathbb{Z})$. Denote by Γ the subsemigroup generated by μ, and by $\mathbb{G} < \mathrm{SL}_d$ the Zariski closure of Γ. Assume*

(a) The measure μ has a finite exponential moment;
(b) The only subspaces of $\mathbb{R}^d$ preserved by Γ are $\{0\}$ and $\mathbb{R}^d$;
(c) The algebraic group $\mathbb{G}$ is Zariski connected.

Let λ_1 denote the top Lyapunov exponent associated to μ. Given $\lambda \in (0, \lambda_1)$, there exists a constant $C = C(\mu, \lambda) > 0$ such that for every Borel probability measure ν on $\mathbb{T}^d$ and every $t \in (0, 1/2)$, if for some $a \in \mathbb{Z}^d \setminus \{0\}$,
$$|\widehat{\mu_n * \nu}(a)| \geq t \quad \text{and} \quad n \geq C \log \frac{\|a\|}{t},$$
then
$$\nu(\mathsf{W}_Q^{(e^{-\lambda n})}) \geq t^C$$
for some $Q \leq \big(\frac{\|a\|}{t}\big)^C$

Proof. Recall the shorthand $\nu_n = \mu_n * \nu$, $n \in \mathbb{N}$. By Proposition 5.2, there is a constant $C_0 > 0$ depending only on μ such that for $m_0 = C_1 \log \frac{\|a\|}{t}$ with $C_1 \geq C_0$, there exists $Q \in [e^{m_0/C_0}, e^{C_0 m_0}]$ such that
$$\nu_{n-m_0}(\mathsf{W}_Q^{(Q^{-8})}) \geq t_0^C.$$

Set $\rho_0 = Q^{-8}$, choose m_1 maximal so that $e^{d\lambda_1 m_1}\rho_0 < Q^{-2}$. Then $e^{d\lambda_1 m_1} \asymp_\mu Q^{-2}\rho_0^{-1} = Q^6$ and hence
$$m_1 \geq \frac{6}{d\lambda_1}\log Q - O_\mu(1) \geq \frac{6C_1}{d\lambda_1 C_0}\log\frac{\|a\|}{t} - O_\mu(1). \tag{5.9}$$

Thus by picking C_1 sufficiently large, we can make $m_1 \geq m_*$ where m_* is the constant given by Proposition 5.5 applied to $\varepsilon := (\lambda_1 - \lambda)/2$.

It is easy to see that if $C = C(\mu, \lambda)$ is chosen large enough, every integer $n \geq C \log \frac{\|a\|}{t}$ can be written as

$$n = m_0 + m_1 + \cdots + m_k$$

for some $k \geq 2$ and some integers $m_2, \ldots, m_k$ satisfying

$$\forall j = 1, \ldots, k-1, \quad m_j < m_{j+1} < \big(1 + \frac{\lambda_1 - \varepsilon}{d\lambda_1}\big) m_j \tag{5.10}$$

Define recursively for $j = 1, \ldots, k$,

$$\rho_j = e^{-(\lambda_1 - \varepsilon) m_j} \rho_{j-1}.$$

Then (5.10) implies, by a simple induction, that

$$\forall j = 1, \ldots, k, \quad e^{d\lambda_1 m_j} \rho_{j-1} < Q^{-2}.$$

Therefore we can apply repeatedly Proposition 5.5 to get

$$\nu(\mathsf{W}_Q^{(\rho_k)}) \geq t^{C_0} - \sum_{j=1}^{k} e^{-\omega m_j}$$

where $\omega > 0$ is a constant depending only on μ and λ. Observe that, first,

$$\rho_k = e^{-(\lambda+\varepsilon)(n - m_0)} \rho_0 \leq e^{-\lambda n}$$

provided that $C \geq \frac{\lambda+\varepsilon}{\varepsilon} C_1$. Secondly, by (5.10),

$$\sum_{j=1}^{k} e^{-\omega m_j} \leq e^{-\omega m_1} \sum_{i \geq 1} e^{-\omega i} \ll_\omega e^{-\omega m_1}$$

is smaller than $t^{C_0}/2$ provided that C_1/C_0 is chosen large enough (recall (5.9)). This finishes the proof. □

6. Conclusion

To conclude this paper, we mention one application of our main theorem, and then give some possible further directions of research, some of which we hope to address in publications to come.

6.1. Expansion in simple groups modulo arbitrary integers. In Section 3, we made use of the result of Salehi Golsefidy and Varjú [38] about expansion in semisimple groups modulo prime – or square-free – numbers. In a reverse direction, it was observed by Bourgain and Varjú [15] that the quantitative equidistibution of linear random walks on the torus of Bourgain, Furman, Lindenstrauss and Mozes [11] could be used to derive some expansion results in $\mathrm{SL}_d(\mathbb{Z}/q\mathbb{Z})$, where q runs over *all* natural integers. Because of the proximality assumption required by [11], their argument could only apply to $\mathbb{R}$-*split* simple $\mathbb{Q}$-groups, such as SL_d. With Theorem 1.2 at hand, we can now generalize their result to any simple $\mathbb{Q}$-group.

Theorem 6.1 (Expansion in simple groups modulo arbitrary integers)**.** *Let S be a finite subset of $\mathrm{GL}_d(\mathbb{Z})$, and Γ the subgroup generated by S. If the Zariski closure of Γ is a simple algebraic group, then the family of Cayley graphs $\mathcal{G}(\pi_q(\Gamma), \pi_q(S))_{q\in\mathbb{N}}$ is a family of expanders .*

As observed by [38], one should expect the theorem to hold with the weaker assumption that the Zariski closure of Γ is perfect. To prove such a result, if one wants to exploit some equidistribution result on the torus similar to Theorem 1.2, one should relax the irreducibility assumption, which leads us to the second point of this conclusion.

6.2. **Without irreducibility.** The only obvious obstruction to equidistribution is when the random walk is trapped in a rational coset of a subtorus that is obtained as the image of a Γ-invariant rational subspace via the projection $\mathbb{R}^d \to \mathbb{T}^d$. Thus, in order to prove equidistribution of a linear random walk on the torus, it may be more natural only to assume the action of Γ to be irreducible on $\mathbb{Q}^d$, rather than $\mathbb{R}^d$.

Indeed, for example, assume that the group generated by the random walk is semisimple and acts strongly irreducibly on $\mathbb{Q}^d$. Then Guivarc'h and Starkov [24] and independently Muchnik [34] showed that every proper closed invariant subset is a finite set of rational points. Moreover, under the same assumption, Benoist and Quint [4, Corollary 1.4] showed that the only non-atomic stationary measure on $\mathbb{T}^d$ is the Haar measure. See also Benoist-Quint [5] for a result on equidistribution of trajectories.

Similarly, Theorem 1.2 should remain valid if one only assumes the irreducibility of the action of Γ on $\mathbb{Q}^d$, as long as the Zariski closure of Γ is semisimple.

The general approach used here should work in this setting, but there is one important difference: the algebra E generated by Γ will no longer be simple, but only semisimple. In particular, the rescaled measure $\tilde{\mu}_n$ studied in Section 3 may very well be concentrated on a proper ideal of E. One therefore needs to modify several of our arguments to adapt the proof to this more general setting.

We hope to address this issue in forthcoming work in collaboration with Elon Lindenstrauss.

Furthermore, the question of equidistribution is still interesting even without any irreducibility assumption. For example, the above-mentioned work of Benoist and Quint gives a classification of orbit closures and stationary measure under the assumption that the Zariski closure of the group is semisimple. In other direction, Bekka and Guivarc'h [2] showed that the action of a subgroup $\Gamma < \mathrm{SL}_d(\mathbb{Z})$ on $\mathbb{T}^d$ has a spectral gap if and only if there is no nontrivial Γ-invariant torus factor on which Γ acts as a virtually abelian group.

6.3. **The two other assumptions.** First, we believe that the Theorem 1.2 is still valid even if one does not require the group $\mathbb{G}$ to be Zariski connected. In fact, many arguments in our proof still works without this assumption, but, as is the case without the irreducibility assumption, the rescaled measures $\tilde{\mu}_n$ may concentrate near a proper subspace of E: the algebra generated by the connected component of G. This leads to several technical difficulties when trying to prove a flattening statement.

Second, it would be an interesting problem to determine what moment conditions are really necessary in order to have the convergence statement of Theorems 1.1 and 1.2. It seems plausible for example that Theorem 1.1 holds with the weaker assumption of a moment of order 1: $\int \log\|g\| \,\mathrm{d}\mu(g) < \infty$. Even a counter-example to Theorem 1.1 without any moment condition would be interesting.

6.4. **Spaces of lattices.** Given the results of Benoist and Quint [4] classifying stationary measures on the space of lattices $\mathrm{SL}_d(\mathbb{R})/\mathrm{SL}_d(\mathbb{Z})$, it is very natural to ask whether one can obtain an analog of Theorem 1.2 in this setting. Even the following qualitative equidistribution problem is still open [3, §5.4. Question 3].

Let μ be a measure on $\mathrm{SL}_d(\mathbb{R})$ generating a Zariski dense subgroup Γ, and x a point in $\mathrm{SL}_d(\mathbb{R})/\mathrm{SL}_d(\mathbb{Z})$ with infinite Γ-orbit. Show that the sequence of measures $(\mu_n * \delta_x)_{n\geq 1}$ converges to the Haar measure as n goes to infinity.

Acknowledgements. We are indebted to Emmanuel Breuillard for several ideas used in §§3.2 and 3.3 and for sharing his unpublished note [17] on non-concentration estimates for random matrix products. It is a pleasure to thank him, as well as Richard Aoun, Yves Benoist, Elon Lindenstrauss, Jean-François Quint and Péter Varjú, for useful and motivating discussions.

References

[1] R. Aoun. Transience of algebraic varieties in linear groups—applications to generic Zariski density. *Ann. Inst. Fourier (Grenoble)*, 63(5):2049–2080, 2013.

[2] B. Bekka and Y. Guivarc'h. On the spectral theory of groups of affine transformations of compact nilmanifolds. *Ann. Sci. Éc. Norm. Supér. (4)*, 48(3):607–645, 2015.

[3] Y. Benoist and J.-F. Quint. Introduction to random walks on homogeneous spaces. *Jpn. J. Math.*, 7(2):135–166, 2012.

[4] Y. Benoist and J.-F. Quint. Stationary measures and invariant subsets of homogeneous spaces (II). *J. Amer. Math. Soc.*, 26(3):659–734, 2013.

[5] Y. Benoist and J.-F. Quint. Stationary measures and invariant subsets of homogeneous spaces (III). *Ann. of Math. (2)*, 178(3):1017–1059, 2013.

[6] Y. Benoist and J.-F. Quint. *Random walks on reductive groups*, volume 62 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics*. Springer, Cham, 2016.

[7] A. Borel. *Linear algebraic groups*, volume 126 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1991.

[8] P. Bougerol and J. Lacroix. *Products of random matrices with applications to Schrödinger operators*, volume 8 of *Progress in Probability and Statistics*. Birkhäuser Boston, Inc., Boston, MA, 1985.

[9] J. Bourgain. Multilinear exponential sums in prime fields under optimal entropy condition on the sources. *Geom. Funct. Anal.*, 18(5):1477–1502, 2009.

[10] J. Bourgain. The discretized sum-product and projection theorems. *J. Anal. Math.*, 112:193–236, 2010.

[11] J. Bourgain, A. Furman, E. Lindenstrauss, and S. Mozes. Stationary measures and equidistribution for orbits of nonabelian semigroups on the torus. *J. Amer. Math. Soc.*, 24(1):231–280, 2011.

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

[13] J. Bourgain and A. Gamburd. Expansion and random walks in $\mathrm{SL}_d(\mathbb{Z}/p^n\mathbb{Z})$. II. *J. Eur. Math. Soc. (JEMS)*, 11(5):1057–1103, 2009. With an appendix by Bourgain.

[14] J. Bourgain and S. V. Konyagin. Estimates for the number of sums and products and for exponential sums over subgroups in fields of prime order. *C. R. Math. Acad. Sci. Paris*, 337(2):75–80, 2003.

[15] J. Bourgain and P. P. Varjú. Expansion in $SL_d(\mathbf{Z}/q\mathbf{Z})$, q arbitrary. *Invent. Math.*, 188(1):151–173, 2012.

[16] R. Boutonnet, A. Ioana, and A. Salehi Golsefidy. Local spectral gap in simple Lie groups and applications. *Invent. Math.*, 208(3):715–802, 2017.

[17] E. Breuillard. A non concentration estimate for random matrix products. Unpublished notes available at https://www.math.u-psud.fr/ breuilla/RandomProducts2.pdf.

[18] E. Breuillard, B. Green, and T. Tao. Approximate subgroups of linear groups. *Geom. Funct. Anal.*, 21(4):774–819, 2011.

[19] D. A. Cox, J. Little, and D. O'Shea. *Ideals, varieties, and algorithms*. Undergraduate Texts in Mathematics. Springer, Cham, fourth edition, 2015. An introduction to computational algebraic geometry and commutative algebra.

[20] N. de Saxcé. A product theorem in simple Lie groups. *Geom. Funct. Anal.*, 25(3):915–941, 2015.

[21] M. D. Fried and M. Jarden. *Field arithmetic*, volume 11 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]*. Springer-Verlag, Berlin, second edition, 2005.

[22] H. Furstenberg. Noncommuting random products. *Trans. Amer. Math. Soc.*, 108:377–428, 1963.

[23] Y. Guivarc'h and A. Raugi. Frontière de Furstenberg, propriétés de contraction et théorèmes de convergence. *Z. Wahrsch. Verw. Gebiete*, 69(2):187–242, 1985.

[24] Y. Guivarc'h and A. N. Starkov. Orbits of linear group actions, random walks on homogeneous spaces and toral automorphisms. *Ergodic Theory Dynam. Systems*, 24(3):767–802, 2004.

[25] W. He. Discretized sum-product estimates in matrix algebras. *ArXiv e-prints 1611.09639*, Nov. 2016. To appear in *Journal d'Analyse Mathématique*.

[26] W. He. Orthogonal projections of discretized sets. *ArXiv e-prints*, page arXiv:1710.00795, Oct. 2017. To appear in Journal of Fractal Geometry.
[27] W. He. *Sums, products and projections of discretized sets.* Phd thesis, Université Paris-Saclay, Sept. 2017.
[28] W. He. Random walks on linear groups satisfying a Schubert condition. *arXiv e-prints*, page arXiv:1905.05695, May 2019. To appear in Israel Journal of Mathematics.
[29] H. A. Helfgott. Growth and generation in $\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$. *Ann. of Math. (2)*, 167(2):601–623, 2008.
[30] S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc. (N.S.)*, 43(4):439–561, 2006.
[31] S. Lang and A. Weil. Number of points of varieties in finite fields. *Amer. J. Math.*, 76:819–827, 1954.
[32] E. Le Page. Théorèmes limites pour les produits de matrices aléatoires. In *Probability measures on groups (Oberwolfach, 1981)*, volume 928 of *Lecture Notes in Math.*, pages 258–303. Springer, Berlin-New York, 1982.
[33] J. Li. Discretized Sum-product and Fourier decay in $\mathbb{R}^n$. *arXiv e-prints*, page arXiv:1811.06852, Nov 2018. arXiv:1811.06852, to appear in Journal d'Analyse Mathématique.
[34] R. Muchnik. Semigroup actions on $\mathbb{T}^n$. *Geom. Dedicata*, 110:1–47, 2005.
[35] M. V. Nori. On subgroups of $\mathrm{GL}_n(\mathbf{F}_p)$. *Invent. Math.*, 88(2):257–275, 1987.
[36] L. Pyber and E. Szabó. Growth in finite simple groups of Lie type. *J. Amer. Math. Soc.*, 29(1):95–146, 2016.
[37] M. S. Raghunathan. *Discrete subgroups of Lie groups.* Springer-Verlag, New York-Heidelberg, 1972. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 68.
[38] A. Salehi Golsefidy and P. P. Varjú. Expansion in perfect groups. *Geom. Funct. Anal.*, 22(6):1832–1891, 2012.
[39] T. Tao. Product set estimates for non-commutative groups. *Combinatorica*, 28(5):547–594, 2008.
[40] T. Tao and V. H. Vu. *Additive combinatorics*, volume 105 of *Cambridge Studies in Advanced Mathematics.* Cambridge University Press, Cambridge, 2010.
[41] B. L. Van der Waerden. *Modern Algebra. Volume II. Based in part on lectures by E. Artin and E. Noether.* Frederick Ungar Publishing Co., New York, transl. from the 3rd german edition, 1950.

Einstein Institute of Mathematics, The Hebrew University of Jerusalem, Jerusalem 91904, Israel.
Email address: `weikun.he@mail.huji.ac.il`

CNRS – Université Paris 13, LAGA, 93430 Villetaneuse, France.
Email address: `desaxce@math.univ-paris13.fr`