arXiv:1510.04043v3 [math.CA] 13 Jun 2018

ENTROPY OF BERNOULLI CONVOLUTIONS AND UNIFORM EXPONENTIAL GROWTH FOR LINEAR GROUPS

EMMANUEL BREUILLARD AND PÉTER P. VARJÚ

ABSTRACT. The exponential growth rate of non polynomially growing subgroups of GL_d is conjectured to admit a uniform lower bound. This is known for non-amenable subgroups, while for amenable subgroups it is known to imply the Lehmer conjecture from number theory. In this note, we show that it is equivalent to the Lehmer conjecture. This is done by establishing a lower bound for the entropy of the random walk on the semi-group generated by the maps $x \mapsto \lambda \cdot x \pm 1$, where λ is an algebraic number. We give a bound in terms of the Mahler measure of λ. We also derive a bound on the dimension of Bernoulli convolutions.

1. INTRODUCTION

A Bernoulli convolution with parameter λ is the distribution μ_λ of the infinite random series $\sum_{n \geqslant 0} \pm \lambda^n$, where the $\pm$ are independent fair coin tosses, and λ is a real number between 0 and 1. Such measures appear in a large number of situations in harmonic analysis and dynamical systems. A key question regarding them, arguably the most puzzling, is asking for which values of λ is this distribution absolutely continuous with respect to Lebesgue measure. If $\lambda < 1/2$, then μ_λ is singular, being supported on a Cantor set. Surprisingly, there is a family of λ's greater than $1/2$ with singular μ_λ (the inverses of the Pisot numbers in $(1, 2)$, see [19]). It is a well-known problem going back to Erdős to determine the set of values of λ for which μ_λ is absolutely continuous, see [19, 40].

If μ_λ is absolutely continuous, then its dimension $\dim \mu_\lambda$ coincides with the dimension of Lebesgue measure, namely 1. In [27], M. Hochman made a breakthrough in this direction, by establishing that $\dim \mu_\lambda = 1$ unless $\lambda \in (\frac{1}{2}, 1)$ is *almost algebraic* in the sense that there is a sequence of degree n polynomials p_n with coefficients in $\{-1, 0, 1\}$ such that $p_n(\lambda)$ tends to 0 super-exponentially fast. It is easily seen that the set of almost algebraic numbers has packing dimension zero, and this was further exploited by Shmerkin [45] to obtain that μ_λ is absolutely continuous for all $\lambda \in (\frac{1}{2}, 1)$, except perhaps for a subset of λ's of Hausdorff dimension zero. (We recall that a set of 0 packing dimension is also of 0 Hausdorff dimension.)

In the first part of this paper we will study the opposite situation when λ is assumed to be an algebraic number. In this case the study of Bernoulli convolutions is closely related to another famous conjecture: the Lehmer conjecture about algebraic numbers. This asserts that the Mahler measure

of an irreducible polynomial in $\mathbb{Z}[X]$ ought to be bounded away from 1 uniformly, unless it is equal to 1.

Recall that if $\pi_\lambda := a_r \prod_1^r (X - \lambda_i) = a_r X^r + \ldots + a_1 X + a_0 \in \mathbb{Z}[X]$ is the minimal polynomial of an algebraic number $\lambda \in \overline{\mathbb{Q}}$, then the Mahler measure M_λ of π_λ is defined by

$$M_\lambda := |a_r| \prod_{|\lambda_i| > 1} |\lambda_i|. \tag{1.1}$$

We can now state our first result:

Theorem 1. *If the Lehmer conjecture holds, then there is $\varepsilon > 0$ such that for every real algebraic number λ with $1 - \varepsilon < \lambda < 1$, the dimension of the Bernoulli convolution μ_λ is 1.*

Moreover, our methods give unconditional results, too. We provide an easily testable sufficient (but unfortunately, not necessary) condition that implies $\dim \mu_\lambda = 1$, which yields plentiful new examples of Bernoulli convolutions with full dimension. See Theorem 5 below, and the discussion that follows it.

In a follow-up paper [9], we prove the following result among others.

Theorem 2. *We have*

$$\{\lambda \in (1/2, 1) : \dim \mu_\lambda < 1\} \subseteq \overline{\{\lambda \in \overline{\mathbb{Q}} \cap (1/2, 1) : \dim \mu_\lambda < 1\}},$$

where $\overline{\mathbb{Q}}$ is the set of algebraic numbers and $\overline{\{\cdot\}}$ denotes the closure of the set with respect to the natural topology of real numbers.

Using this result, we can drop the condition of algebraicity from Theorem 1. If Lehmer's conjecture holds, then there is $\varepsilon > 0$ such that $\dim \mu_\lambda = 1$ for all $1 - \varepsilon < \lambda < 1$. In addition, Theorem 2 provides additional motivation for studying the case of algebraic parameters. Indeed, if one was able to prove that the inverses of Pisot numbers are the only algebraic parameters such that $\dim \mu_\lambda < 1$, then we would have $\dim \mu_\lambda = 1$ for all transcendental parameters, since the set of Pisot numbers is closed (see [44]). Unfortunately, we are not able to prove this, but we believe that our methods introduced in this paper may yield an approach.

In the second part of this paper, we discuss a connection between Bernoulli convolutions and a classical topic in geometric group theory, namely the growth of finitely generated groups. Given a group Γ generated by a finite subset S, we denote by ρ_S the rate of exponential growth of its n-th powers, namely:

$$\rho_S := \lim_{n \to +\infty} \frac{1}{n} \log |S^n|,$$

where S^n is the set of products of n elements chosen from S.

Examples of Grigorchuk and de la Harpe [24] show that ρ_S can take arbitrarily small positive values even for linear groups. However in [10, Conjecture 1.1], the first author made the following conjecture:

Conjecture 3 (Growth Conjecture). *Given $d \in \mathbb{N}$, there is $\varepsilon = \varepsilon(d) > 0$ such that for every finite subset S in $\mathrm{GL}_d(\mathbb{C})$, either $\rho_S = 0$, or $\rho_S > \varepsilon$.*

A positive answer to this conjecture was obtained earlier in [11,14] in the special case when S generates a non-virtually solvable subgroup of $\mathrm{GL}_d(\mathbb{C})$. Indeed this is a simple consequence of the uniform version of the Tits alternative proved therein (see also [8, 20] for earlier related works). A key ingredient in this work was the proof [13] of an analogue of Lehmer's conjecture in the setting of semisimple algebraic groups.

The case when the subgroup generated by S is virtually solvable, i.e. contains a solvable subgroup of finite index, is surprisingly harder, since it was observed in [12] that the above conjecture, already in the case of solvable subgroups of GL_2, implies the Lehmer conjecture.

We can now show the converse:

Theorem 4. *The Growth conjecture is equivalent to the Lehmer conjecture.*

The key behind the proofs of Theorems 1 and 4 is the study of the entropy of Bernoulli convolutions with algebraic parameter via Theorem 5 below, which directly relates the Mahler measure to the entropy.

The ping-pong method used in most proofs of exponential word growth is not powerful enough in our situation. Indeed no free semi-group may in general be generated by words of small length (see [12, Thm 1.7]). Fortunately here entropy comes to our rescue.

While it is clear how entropy relates to an exponential growth rate (Jensen's inequality, see (1.5)), the relation between entropy and the dimension of a Bernoulli convolution is provided by Hochman's theorem alluded above, which tells us that if $\lambda \in \overline{\mathbb{Q}} \cap (1/2, 1)$, then

$$\dim \mu_\lambda = \min \left(\frac{h_\lambda}{|\log \lambda|}, 1 \right), \tag{1.2}$$

where h_λ is the *entropy* of the random walk on the semi-group generated by the affine transformations $x \mapsto \lambda \cdot x \pm 1$. More precisely, it is defined as follows:

$$h_\lambda := \lim_{n \to +\infty} \frac{H(\mu_\lambda^{(n)})}{n}, \tag{1.3}$$

where $H(\theta)$ denotes the Shannon entropy of the discrete probability measure θ and $\mu_\lambda^{(n)}$ is the law of the random variable $\mu_\lambda^{(n)} = \sum_{i=0}^{n-1} \pm \lambda^i$, where $\pm$ are independent fair coin tosses. If $X_1, \ldots, X_n$ are independent random similarities, which take the values $x \mapsto \lambda \cdot x + 1$ or $x \mapsto \lambda \cdot x - 1$ with equal probability, then $\mu_\lambda^{(n)}$ is also the law of $X_n \cdots X_1(0)$. This explains our terminology of calling h_λ the entropy of the random walk. We explain in Paragraph 3.4 how (1.2) follows from [27].

It is convenient for us to use the following convention for the base of logarithms. *We write* log *for the base* 2 *logarithm, so that* $\log 2 = 1$. We can now state our main theorem.

Theorem 5. *There is a positive constant* $c > 0$, *such that the following holds. Let* λ *be an algebraic number. Then*

$$c \min\{1, \log M_\lambda\} \leqslant h_\lambda \leqslant \min\{1, \log M_\lambda\}.$$

We stress that λ here can take any complex value in $\overline{\mathbb{Q}}$, not only real values. Our (non-rigorous) numerical calculations show that one can take $c = 0.44$ in the above result. This is probably not optimal.

Remark 6. In fact, the result holds in greater generality. Let $\xi_0, \xi_1, \ldots$ be a sequence of finitely supported i.i.d. random variables with common law ν_0. Suppose that $\xi_n \in \mathbb{Q}$ almost surely. Write $\mu_{\lambda,\nu_0}^{(n)}$ for the law of the random variable $\sum_{i=0}^{n-1} \xi_i \lambda^i$ and write h_{λ,ν_0} for the quantity that we obtain when we replace $\mu_\lambda^{(n)}$ by $\mu_{\lambda,\nu_0}^{(n)}$ in (1.3). Then there is a constant $c(\nu_0)$ depending only on ν_0 such that

$$c(\nu_0)\min\{1, \log M_\lambda\} \leq h_{\lambda,\nu_0} \leq \min\{H(\nu_0), \log M_\lambda\}.$$

Moreover, for each $\varepsilon > 0$, there is a measure ν_0 such that $c(\nu_0) > 1 - \varepsilon$. This means, in particular, that for every algebraic $\lambda \in (1/2, 1)$ such that λ^{-1} is neither a Pisot nor a Salem number, there is a measure ν_0 supported on the integers, such that $\dim \mu_{\lambda,\nu_0} = 1$. (Without the requirement that ν_0 is supported on the integers, this has been known before.)

Our main motivation for considering this more general case is an application in the paper [48] for the absolute continuity of certain biased Bernoulli convolutions.

The upper bound in Theorem 5 is often a strict inequality. Indeed we prove:

Proposition 7. *Let λ be an algebraic number such that $M_\lambda < 2$ and assume that λ has no Galois conjugates on the unit circle. Then $h_\lambda < \log M_\lambda$.*

We note that this proposition was proved by Garsia [23] in the case where λ is the inverse of a Pisot number in $(1, 2)$.

The upper bound in Theorem 5 follows from a simple counting argument. In fact if we denote by ρ_λ the rate of exponential growth (in base 2) of the size of the support of $\mu_\lambda^{(n)}$, namely

$$\rho_\lambda := \lim_{n \to +\infty} \frac{\log |\operatorname{Supp}(\mu_\lambda^{(n)})|}{n}, \tag{1.4}$$

then an obvious upper bound on h_λ is

$$h_\lambda \leqslant \rho_\lambda \leqslant \log 2 = 1, \tag{1.5}$$

as follows say from Jensen's inequality, see (2.9).

It is easy to see that $\rho_\lambda = 1$ unless λ is a root of a polynomial with coefficients in $\{-1, 0, 1\}$. For a topological study of the set of roots of such polynomials, see [3].

If λ is algebraic, then one has the following easy upper bound (see Lemma 16)

$$\rho_\lambda \leqslant \log M_\lambda, \tag{1.6}$$

from which it also follows that $M_\lambda \geqslant 2$ unless λ is a root of a polynomial with $\{-1, 0, 1\}$ coefficients, a fact known since [38, 39].

For the strict inequality in Proposition 7, one needs a slightly more precise upper bound on the size of the support of $\mu_\lambda^{(\ell)}$, which is obtained by first showing that the law of the image of $\mu_\lambda^{(\ell)}$ under the geometric embedding $\mathbb{Q}(\lambda) \to \mathbb{R}^{n+2m}$ (where n is the number of real embeddings, $2m$ the number of complex embeddings, and λ has modulus less than 1) is singular with

respect to Lebesgue measure. It is an adaptation of Garsia's proof [23] of this proposition in the special case when λ^{-1} is Pisot.

Theorem 1 is a direct corollary of Theorem 5 and Hochman's identity (1.2). We note that Theorem 5 also has unconditional consequences, since Lehmer's conjecture has been verified for several classes of algebraic numbers, see e.g. [2, 6]. Note that already Hochman's identity readily implies that $\dim \mu_\lambda = 1$, provided $\lambda \in \overline{\mathbb{Q}} \cap (\frac{1}{2}, 1)$ and the semi-group generated by $x \mapsto \lambda \cdot x \pm 1$ is free, e.g. when λ is not a unit or has a conjugate outside the annulus $1/2 \leq |z| \leq 2$. Theorem 5 provides another condition implying $\dim \mu_\lambda = 1$, which can be easily tested, namely:

Corollary 8. *If λ is a real algebraic number such that*

$$\min(M_\lambda, 2)^{-0.44} \leqslant \lambda \leqslant 1,$$

then $\dim \mu_\lambda = 1$.

Here we used are numerical estimate $c_0 \geq 0.44$ on the constant in Theorem 5.

The derivation of Theorem 4 from Theorem 5 is based on some group theoretic arguments, which enable one to give a lower bound on the growth rate of an arbitrary virtually solvable subgroup of $\mathrm{GL}_d(\mathbb{C})$ in terms of the growth rate of the semi-group generated by the two affine transformations of the complex line $x \mapsto \lambda x + 1$ and $x \mapsto \lambda x - 1$, which is precisely ρ_λ. See Theorem 20 in Section 4.

The proof of the lower bound in Theorem 5 is the main contribution of this paper. Our argument is a multi-scale analysis that exploits the self-similarity (see (2.2) below) of the measure μ_λ, together with an inequality for the entropy of the sum of independent random variables (see Proposition 10), which can be seen as an entropy analogue of the Plünnecke-Ruzsa inequality from additive combinatorics and is discussed in [31, 35, 46]. This allows to lower bound the entropy by a sum of entropy contributions at each scale, each of which is uniformly bounded below.

Finally, we mention some related works from the literature.

In [26] Hare and Sidorov computed explicit lower bounds for the entropy h_λ in the special case, when λ is a Pisot number. In particular they showed, that $h_\lambda > 0.81 \cdot \log \lambda^{-1}$ for all Pisot numbers, which is a better constant than what our methods yield. Moreover, Sidorov pointed out to us that their result holds in greater generality for all algebraic numbers. On the other hand, Theorem 5 provides an estimate for h_λ in terms of the Mahler measure with a constant independent of the number of Galois conjugates inside the unit disk. This uniformity is the main difficulty and the main point in our work. See also [1] for upper and lower bounds on h_λ, in the case when λ is the Golden Ratio.

Paul Mercat studied the quantity ρ_λ in his thesis [37]. He showed that $\rho_\lambda = \log M_\lambda$, whenever $\lambda \in (1/2, 1)$ is a Salem number. This is also true for Pisot numbers [34]. Mercat also showed that when λ has no conjugates on the unit circle, then the semi-group generated by $x \to \lambda x \pm 1$ is automatic, in particular 2^{ρ_l} is an algebraic number in this case, and he gave an algorithm to compute its minimal polynomial.

The following result of Peters [41] was brought to our attention by Andreas Thom. For each integer N and algebraic number λ, denote by $S_{\lambda,N}$ the set of matrices

$$\begin{pmatrix} \lambda & n \\ 0 & 1 \end{pmatrix},$$

where n ranges through the integers $-N, \dots, N$. It is proved in [41] that

$$\lim_{N \to \infty} \rho_{S_{\lambda,N}} = \log M_\lambda.$$

The main interest in our results is that we are able to give a good lower bound even for $\rho_{S_{\lambda,1}}$.

The paper is organized as follows. In Section 2 we outline the proof of the lower bound in Theorem 5, recall the basic properties of entropy and prove some entropy inequalities we need. Section 3 completes the proof of the lower bound in Theorem 5 and also deals with the upper bound estimates. In Paragraph 3.4 we explain how (1.2) follows from [27] and deduce Theorem 1. In Section 4 we discuss the applications to the growth of solvable linear groups and prove Theorem 4. Finally the last section is devoted to some open problems.

Acknowledgements. It is a pleasure to thank Francesco Amoroso, Jean Bourgain, Mike Hochman, Paul Mercat, Nikita Sidorov and Andreas Thom for interesting discussions. We are grateful to the referee for valuable comments and suggestions, which improved the paper.

2. Entropy bounds for the random walk

In this section, we set notations, review some basic properties of entropy, give an outline of the proof of Theorem 5 and discuss a number of preliminaries.

2.1. **Bernoulli convolutions in matrices.** Let $A \in M_d(\mathbb{R})$ be a matrix with real entries and spectral radius strictly less than 1. We fix an atomic probability measure ν_0 supported on a finite subset of the rationals. Let $\{\xi_n\}_{n \geqslant 0}$ be a sequence of bounded i.i.d. random variables with common law ν_0. The power series $\sum_{n \geqslant 0} A^n$ converges absolutely, and hence the random variable

$$X_A := \sum_{n \geqslant 0} \xi_n A^n \tag{2.1}$$

is well-defined and almost surely finite. We denote its law by μ_A. Its support is some bounded region of $M_d(\mathbb{R})$, and it satisfies the self-similarity relation

$$\mu_A = \mu_A^{(\ell)} * A^\ell \mu_A, \tag{2.2}$$

where $\mu_A^{(\ell)}$ is the law of the random variable

$$X_A^{(\ell)} := \sum_{0 \leqslant n \leqslant \ell-1} \xi_n A^n,$$

and $A^\ell \mu_A$ is the push-forward of the measure μ_A by the linear map $x \mapsto A^\ell x$. If $A = \lambda \in M_1(\mathbb{R})$ then this is consistent with the notation μ_λ and $\mu_\lambda^{(\ell)}$ used in the Introduction. (For brevity we omit the subscript ν_0 in our notation.)

For any given vector $x \in \mathbb{R}^d$, we let $X_{A,x} := X_A \cdot x$ and $X_{A,x}^{(\ell)} := X_A^{(\ell)} \cdot x$ and denote the associated measures on $\mathbb{R}^d$ by $\mu_{A,x}$ and $\mu_{A,x}^{(\ell)}$. Similarly to μ_A, this measure satisfies the relation

$$\mu_{A,x} = \mu_{A,x}^{(\ell)} * A^\ell \mu_{A,x}. \tag{2.3}$$

2.2. **Outline of the proof of Theorem 5.** In this paper $H(X)$ denotes the entropy of the random variable X taking values in $\mathbb{R}^d$. More precisely, if X takes only countably many values $\{x_i\}_i$, then $H(X)$ will denote the *Shannon entropy*, i.e.

$$H(X) = -\sum_i p_i \log p_i,$$

where p_i is the probability that $X = x_i$. If on the other hand the distribution of X is absolutely continuous with respect to Lebesgue measure, then $H(X)$ will denote the *differential entropy*, that is

$$H(X) = -\int f(x) \log f(x) dx,$$

where $f(x)$ is the density of X, which is well defined provided $f \log f$ is in $\mathbb{L}^1(\mathbb{R}^d)$. Although we will use the same letter for the Shannon and the differential entropy, it should not cause confusion: we will only consider the entropy of random variables whose law is either atomic, and we will then use the Shannon entropy, or absolutely continuous with respect to Lebesgue, in which case the differential entropy will be used.

We fix an algebraic number $\lambda \in \overline{\mathbb{Q}}$ and assume that $A \in M_d(\mathbb{R})$ is such a matrix whose eigenvalues coincide with the Galois conjugates λ_i of λ of modulus < 1. Note that the spectral radius of such an A is strictly less than 1. Of course it is always possible to find such a matrix, because the non-real Galois conjugates, i.e. the roots of the minimal polynomial π_λ, come in pairs of complex conjugates.

We make the following simple observation: if λ satisfies an equation of the form $\sum_0^{\ell-1} \varepsilon_n \lambda^n = \sum_0^{\ell-1} \varepsilon_n' \lambda^n$ for two sets of rationals $\{\varepsilon_n\}_{0 \leqslant n \leqslant \ell-1}$ and $\{\varepsilon_n'\}_{0 \leqslant n \leqslant \ell-1}$, then every Galois conjugate of λ will satisfy the same equation, and hence (looking at a basis of eigenvectors of A) we also have for all $x \in \mathbb{R}^d$

$$\sum_{n=0}^{\ell-1} \varepsilon_n A^n x = \sum_{n=0}^{\ell-1} \varepsilon_n' A^n x.$$

From this we can deduce that:

$$H(X_\lambda^{(\ell)}) = H(X_A^{(\ell)}) \geqslant H(X_{A,x}^{(\ell)}). \tag{2.4}$$

Our goal (towards Theorem 5) is to obtain a lower bound on $H(X_{A,x}^{(\ell)})$. We shall shortly describe a method to do this, but first introduce some more notation. For a matrix $B \in \mathrm{GL}_d(\mathbb{R})$ and a bounded random variable X on $\mathbb{R}^d$, $H(X; B)$ will denote the quantity

$$H(X; B) := H(X + G_B) - H(G_B),$$

where G_B is a centered gaussian random variable with co-variance matrix $B\,{}^t B$ that is independent of X. (Here ${}^t B$ denotes the transpose of B.) If

$B_1, B_2 \in M_d(\mathbb{R})$ are two matrices, then $H(X; B_1|B_2)$ will denote

$$H(X; B_1|B_2) := H(X; B_1) - H(X; B_2).$$

These quantities have the following intuitive meaning. Denote by Δ the unit ball in $\mathbb{R}^d$. Then $H(X; B)$ measures the amount of information needed to describe the law of X up to an error inside $B(\Delta)$. One may define similar quantities using other mollifiers than the standard gaussian (say, the indicator function of Δ or some other bump like function), which would capture the same intuitive meaning. These quantities differ from $H(X; B)$ by an additive constant depending on the dimension. However, our arguments cannot tolerate any losses, so the choice of the smoothing is important for technical reasons.

We use the following partial order on $M_d(\mathbb{R})$. We write $B_1 \preceq B_2$ if $B_2{}^t B_2 - B_1{}^t B_1$ is a non-negative semi-definite matrix, or equivalently $\|{}^t B_1 x\|_2 \leq \|{}^t B_2 x\|_2$ for all $x \in \mathbb{R}^d$. The quantities defined above enjoy the following properties, which will be crucial for us.

Lemma 9. *Let $B_1, B_2 \in \mathrm{GL}_d(\mathbb{R})$ be such that $B_1 \preceq B_2$. Assume that X, Y are two bounded independent random variables taking values in $\mathbb{R}^d$. Then*

   (i) $H(X; B_1) \geqslant 0$,
   (ii) $H(X; B_1) + H(Y; B_1) \geqslant H(X + Y; B_1)$,
   (iii) $H(X; B_1) \geqslant H(X; B_2)$,
   (iv) $H(X + Y; B_1|B_2) \geqslant H(X; B_1|B_2)$.

This lemma will be proved in Section 2.4. For now we pursue our outline of the proof of Theorem 5. Recall that we started with a matrix $A \in M_d(\mathbb{R})$ whose eigenvalues coincide with the Galois conjugates of λ of modulus strictly less than 1. Assume now that its operator norm (for the canonical Euclidean structure on $\mathbb{R}^d$) is less or equal to 1. This ensures that

$$A^{i+1} \preceq A^i$$

for each $i \in \mathbb{N}$.

The first step in the proof will be to approximate $H(X_{A,x}^{(\ell)})$ by $H(X_{A,x}; A^\ell|\mathrm{Id})$, where Id denotes the identity matrix. Intuitively this is expected to hold, since $X_{A,x}$ differs from $X_{A,x}^{(\ell)}$ only at a scale proportional to $A^\ell(\Delta)$. In fact, we will show that even $H(X_{A,x}^{(\ell)}) \geqslant H(X_{A,x}; A^\ell|\mathrm{Id})$ holds.

Then we write

$$H(X_{A,x}; A^\ell|\mathrm{Id}) = \sum_{i=1}^{\ell} H(X_{A,x}; A^i|A^{i-1}).$$

Using Lemma 9 (iv) and the self-similarity property (2.3), we will bound from below each term on the right hand side by $H(X_{A,x}; A|\mathrm{Id})$. Taking the limit $\ell \to \infty$ we then obtain

$$h_{\lambda,\nu_0} = \lim \frac{H(X_A^{(\ell)})}{\ell} \geqslant H(X_{A,x}; A|\mathrm{Id}). \tag{2.5}$$

To estimate the right hand side of (2.5), we use inequality (iv) from Lemma 9 and keep only one convolution factor in $X_{A,x}$. This yields the

following inequality, which is valid for all vectors $x \in \mathbb{R}^d$, and all real matrices A with norm at most 1 and eigenvalues equal to the conjugates of λ of modulus less than 1,

$$h_{\lambda,\nu_0} \geqslant H(\xi_0 \cdot x; A|\mathrm{Id}). \tag{2.6}$$

By a suitable choice of the vector x and the matrix A we reduce this to entropies on $\mathbb{R}$:

$$h_{\lambda,\nu_0} \geqslant \Phi(M_\lambda) := \sup_{t>0}\{H(tM_\lambda\xi_0 + G) - H(t\xi_0 + G)\}, \tag{2.7}$$

where G is a standard gaussian random variable independent of ξ_0.

Then a calculus exercise allows to get the lower bound $c(\nu_0)\min\{1, \log M_\lambda\}$ for the right hand side above, thus concluding our outline of the proof of Theorem 5.

In the remainder of this section, we recall some basic facts about Shannon and differential entropies in Paragraph 2.3, then prove Lemma 9 in Paragraph 2.4. The proof of Theorem 5 is given in the next section following the above outline.

2.3. **Basic properties of Entropy.** Recall that we denote by $H(X)$ the Shannon entropy of X if X is a discrete random variable in $\mathbb{R}^d$ and the differential entropy if X is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$. We refer the reader to [16] for a thorough introduction to information theory and entropy. The purpose of this paragraph is to recall a few properties.

The Shannon entropy is always non-negative. The differential entropy on the other hand can take negative values. For example, if $A \in \mathrm{GL}_d(\mathbb{R})$, and X is a random variable with finite differential entropy $H(X)$, then it follows from the change of variables formula that

$$H(AX) = H(X) + \log|\det A|, \tag{2.8}$$

which can take negative values when A varies. On the other hand, if X takes countably many values, the Shannon entropy of AX is the same as that of X. Note that both entropies are invariant under translation by a constant in $\mathbb{R}^d$.

The density of a centered gaussian random variable G_A with co-variance matrix $A\,^tA$ on $\mathbb{R}^d$ is

$$g_A(x) := \frac{1}{(2\pi)^{d/2}|\det A|} \cdot e^{-||A^{-1}x||^2/2},$$

and its differential entropy is $\frac{d}{2}\log 2e\pi + \log|\det A|$. It maximizes the differential entropy of a random variable in $\mathbb{R}^d$ with the same co-variance. For a proof see [16, Example 12.2.8].

We define $F(x) := -x\log(x)$ for $x > 0$ and recall that F is concave, and it is sub-additive, i.e. $F(x+y) \leqslant F(x) + F(y)$, and it also satisfies the identity $F(xy) = xF(y) + yF(x)$.

From the concavity of F and Jensen's inequality, we see that for any atomic random variable X taking at most N possible different values,

$$H(X) \leqslant \log N. \tag{2.9}$$

Let now X and Y be two independent random variables in $\mathbb{R}^d$. If both are atomic, it follows immediately from the sub-additivity of $F(x)$ and the identity $F(xy) = xF(y) + yF(x)$ that $H(X+Y) \leqslant H(X) + H(Y)$ for Shannon entropy. This is no longer true for differential entropy (since the formula is not invariant under a linear change of variable). However if X is atomic and bounded, while Y is assumed absolutely continuous, then

$$H(X + Y) \leqslant H(X) + H(Y), \tag{2.10}$$

where $H(X)$ is Shannon's entropy and the other two are differential entropies. To see this, note that if $f(y)$ is the density of Y, then the density of $X + Y$ is $\mathbb{E}(f(y - X)) = \sum_i p_i f(y - x_i)$, hence:

$$
\begin{aligned}
H(X + Y) &= \int F\Big(\sum_i p_i f(y - x_i)\Big) dy \\
&\leqslant \int \sum_i F(p_i f(y - x_i)) dy \\
&= \int \sum_i F(p_i) f(y - x_i) dy + \int \sum_i p_i F(f(y - x_i)) dy \\
&= \sum_i F(p_i) + \int F(f(y)) dy = H(X) + H(Y)
\end{aligned}
$$

and (2.10) follows.

In the other direction, we always have the lower bound

$$H(X + Y) \geqslant \max\{H(X), H(Y)\} \tag{2.11}$$

if all three entropies are of the same type (i.e. either Shannon or differential), as follows easily from the concavity of F.

The *relative entropy* (or Kullback-Leibler divergence) of two absolutely continuous probability distributions p and q on $\mathbb{R}^d$ is defined as

$$D(p||q) := \int p(x) \log \frac{p(x)}{q(x)} dx$$

This quantity is always non-negative (*information inequality*) as follows immediately from Jensen's inequality and the concavity of $x \mapsto \log x$, since

$$-D(p||q) = \int p(x) \log \frac{q(x)}{p(x)} dx \leqslant \log \Big(\int p(x) \frac{q(x)}{p(x)} dx \Big) = 0.$$

Moreover $D(p||q) = 0$ if and only if p and q coincide almost everywhere.

A direct consequence of this inequality is the fact that the entropy of the joint law of two random variables (X, Y) is at most the sum of the entropy of each marginal, namely:

$$H(X, Y) \leqslant H(X) + H(Y)$$

Indeed the difference $H(X) + H(Y) - H(X, Y)$ (also called the *mutual information*) can be expressed as the relative entropy

$$D(p_{(X,Y)}(x, y)||p_X(x)p_Y(y)),$$

where $p_{(X,Y)}(x, y)$ is the joint density of (X, Y) and $p_X(x)$ and $p_Y(y)$ the marginals, i.e. the densities of X and Y respectively. Equality holds if and only if X and Y are independent.

A similar inequality is as follows. Suppose X, Y, Z are three $\mathbb{R}^d$-valued absolutely continuous random variables such that the triple (X, Y, Z) as well as the individual marginals X, Y and Z have finite differential entropy. Then

$$H(X, Y, Z) + H(Z) \leqslant H(X, Z) + H(Y, Z) \qquad (2.12)$$

Indeed, the difference $H(X, Z) + H(Y, Z) - H(X, Y, Z) - H(Z)$ is exactly the relative entropy $D(p||q)$ of the two $(\mathbb{R}^d)^3$-valued probability distributions $p(x, y, z) := p_{(X,Y,Z)}(x, y, z)$ and $q(x, y, z) := p_{(X,Z)}(x, z) p_{(Y,Z)}(y, z) / p_Z(z)$.

2.4. **Inequalities for entropies of sums of random variables.** The purpose of this section is to prove Lemma 9. We first recall the following result from [35, Theorem I.].

Theorem 10 (submodularity inequality). *Assume that X, Y, Z are three independent $\mathbb{R}^d$-valued random variables such that the distributions of Y, $X + Y$, $Y + Z$ and $X + Y + Z$ are absolutely continuous with respect to Lebesgue measure and have finite differential entropy. Then*

$$H(X + Y + Z) + H(Y) \leqslant H(X + Y) + H(Y + Z). \qquad (2.13)$$

This result goes back in some form at least to a paper by Kaimanovich and Vershik [33, Proposition 1.3], which related the positivity of the entropy of a random walk on a group to the existence of bounded harmonic functions. The version in that paper assumes that the laws of X, Y and Z are identical. The inequality was rediscovered by Madiman [35, Theorem I.] in the greater generality stated above. Then it was recast in the context of entropy analogues of sumset estimates from additive combinatorics by Tao [46] and Kontoyannis and Madiman [31]. And indeed Theorem 10 can be seen as an entropy analogue of the *Plünnecke–Ruzsa inequality* in additive combinatorics (see [47, Corollary 6.29]).

We provide the short proof for the reader's convenience. We first give it under the additional assumption that X and Z are absolutely continuous and have finite differential entropy.

Proof of Theorem 10 assuming that X, Z have finite differential entropy. We apply (2.12) to the random variables $X' = X, Y' = X + Y, Z' = X + Y + Z$ to get

$$H(X, X+Y, X+Y+Z) + H(X+Y+Z) \leqslant H(X, X+Y+Z) + H(X+Y, X+Y+Z)$$

However $H(X, X + Y, X + Y + Z) = H(X, Y, Z)$ because the linear transformation used here has determinant 1 (see (2.8)), while $H(X, X + Y + Z) = H(X, Y + Z) = H(X) + H(Y + Z)$ by independence of X, Y, Z. Similarly $H(X + Y, X + Y + Z) = H(X + Y) + H(Z)$, and $H(X, Y, Z) = H(X) + H(Y) + H(Z)$. The result follows. $\square$

For the general case we need to approximate X and Z by absolutely continuous random variables. We will replace them by $X + \varepsilon G_{\mathrm{Id}}$ and $Z + \varepsilon G'_{\mathrm{Id}}$, where G_{Id} and G'_{Id} are two independent (of everything) gaussian random variables with covariance matrices Id. We will need the following two simple Lemmata.

Lemma 11. *Let X be an absolutely continuous bounded random variable in $\mathbb{R}^d$ with finite differential entropy. Let G_{Id} be a standard gaussian random variable in $\mathbb{R}^d$ that is independent of X. Then $H(X) = \lim_{\varepsilon \to 0} H(X + \varepsilon G_{\mathrm{Id}})$*

Proof. We have $H(X) \leq H(X + \varepsilon G_{\mathrm{Id}})$ by (2.11) for all ε.

Denote by f the density of X and by g_ε the density of $\varepsilon G_{\mathrm{Id}}$. We observe that $f * g_\varepsilon$ converges to f almost everywhere and we set out to construct a majorant function. We fix a sufficiently large real number R such that $f(x) = 0$ for all $\|x\| > R - 10$. We observe that for all x with $\|x\| > R$ and $0 < \varepsilon < 1$ we have $f * g_\varepsilon(x) < e^{-(\|x\|-R)^2}$. We write m for the maximum of the function $F(x) = -x \log(x)$ and define $M(x) = m$ for $\|x\| \leq R + 10$ and $M(x) = (\|x\| - R)^2 e^{-(\|x\|-R)^2}$ for $\|x\| > R + 10$.

We observe that $F(f * g_\varepsilon(x)) \leq M(x)$ for all x, hence by Fatou's Lemma we have

$$H(X) \geq \limsup_{\varepsilon \to 0} \int F(f * g_\varepsilon(x)) dx = \limsup_{\varepsilon \to 0} H(X + \varepsilon G_{\mathrm{Id}}).$$

This completes the proof. $\qquad\qquad\square$

If X is a random variable and A is an event with positive probability, then we denote by $X|_A$ a random variable that satisfies $\mathbb{P}(X|_A \in B) = \mathbb{P}(X \in B \cap A)/\mathbb{P}(A)$ for any measurable subset B of the domain of X.

Lemma 12. *Let X be an absolutely continuous random variable in $\mathbb{R}^d$ with finite differential entropy. Let $\{A_n\}$ be an increasing sequence of events such that*

$$\lim_{n \to \infty} \mathbb{P}(A_n) = 1.$$

Then $H(X) = \lim_{n\to\infty} H(X|_{A_n})$.

Proof. Let M be a monotone increasing function on $\mathbb{R}_{>0}$ such that $M(x) = |x \log(x)|$ if $x \leq 1/10$ or $x \geq 10$ and $M(x) \geq |x \log(x)|$ for all x. We write f for the density of X and $\mathbb{P}(A_n)^{-1} f_n$ for the density of $X|_{A_n}$. We note that f_n converges to f almost everywhere and that $|f_n \log(f_n)| \leq M(f)$. Since $M(x) \leq Cx + |x \log x|$ for some constant C and both f and $|f \log(f)|$ are in L^1, we see that $M(f) \in L^1$. Thus by the dominated convergence theorem, we have

$$H(X) = \lim_{n \to \infty} -\int f_n(x) \log f_n(x) dx = \lim_{n \to \infty} H(X|_{A_n}).$$

$$\qquad\qquad\qquad\square$$

Proof of Theorem 10 in the general case. First we assume that X, Y and Z are all bounded. Let G_{Id} and G'_{Id} be two independent (from each other and X, Y, Z) standard gaussian random variables of dimension matching that of X. We apply the already proved case of the theorem for the random variables $X + \varepsilon G_{\mathrm{Id}}$, Y and $Z + \varepsilon G'_{\mathrm{Id}}$ and obtain

$$H(X+Y+Z+\varepsilon G_{\mathrm{Id}}+\varepsilon G'_{\mathrm{Id}}) + H(Y) \leqslant H(X+Y+\varepsilon G_{\mathrm{Id}}) + H(Y+Z+\varepsilon G'_{\mathrm{Id}}).$$

In light of Lemma 11, letting $\varepsilon \to 0$ we can conclude the theorem for X, Y, Z.

If any of X, Y or Z is unbounded, then we define A_n to be the event $\|X\| + \|Y\| + \|Z\| < n$. Using the theorem for the bounded variables $X|_{A_n}$, $Y|_{A_n}$ and $Z|_{A_n}$ we get

$$H([X + Y + Z]|_{A_n}) + H(Y|_{A_n}) \leqslant H([X + Y]|_{A_n}) + H([Y + Z]|_{A_n}).$$

We take the limit $n \to \infty$ and conclude the theorem from Lemma 12. $\qquad \square$

Recall the definition of $H(X; B)$ and $H(X; B_1|B_2)$ from the two paragraphs before Lemma 9. We finish this section by proving the properties of these entropies claimed in this lemma.

Proof of Lemma 9. Item (i) follows easily from the concavity of $F(x) = -x\log(x)$, indeed

$$H(X + G_{B_1}) = \int F(\mathbb{E}(g_{B_1}(x - X)))dx \geqslant \int \mathbb{E}(F(g_{B_1}(x - X)))dx = H(G_{B_1}).$$

Item (ii) is a consequence of (2.13) applied to the three independent variables $X' = X$, $Y' = G_{B_1}$ and $Z' = Y$.

Since $B_1 \preceq B_2$, there exists $M \in M_d(\mathbb{R})$ such that $B_2 {}^tB_2 = B_1 {}^tB_1 + M {}^tM$. In particular if G_{B_1} and G_M are two independent centered gaussian distributions on $\mathbb{R}^d$ with co-variance matrix $B_1 {}^tB_1$ and $M {}^tM$ respectively, then $G_{B_1} + G_M$ is a centered gaussian variable with co-variance $B_2 {}^tB_2$.

Now item (iii) follows from (2.13) by setting $Y = G_{B_1}$ and $Z = G_M$, while item (iv) also follows from (2.13) applied to the independent variables X', Y', Z' defined by $X' := Y$, $Y' := X + G_{B_1}$, $Z' := G_M$. $\qquad \square$

3. Proof of Theorem 5 and Proposition 7

In this section we establish Theorem 5 and Proposition 7. In Paragraph 3.1 we give the details of the above outline and give a lower bound for h_λ in terms of M_λ only, via the function Φ defined above in (2.7). In Paragraph 3.2 we study the function Φ and deduce the desired lower bound in Theorem 5. Finally, we prove the upper bounds of Theorem 5 and Proposition 7 in Paragraph 3.3.

3.1. **Lower bounds on the entropy.** We keep the notations introduced in the previous section. In particular ξ_0 is a random variable with law ν_0, and G is an independent standard gaussian real random variable. Also M_λ is the Mahler measure of the algebraic number λ. We recall that for $a > 0$

$$\Phi_{\nu_0}(a) = \sup_{t > 0}\{H(\xi_0 ta + G) - H(\xi_0 t + G)\}.$$

The following proposition is the main goal of this paragraph. It establishes the lower bounds (2.5)–(2.7) from our outline.

Proposition 13. *Let λ be an algebraic number and $A \in M_d(\mathbb{R})$ a matrix such that $\|A\| \leq 1$ and the eigenvalues of A coincide with the Galois conjugates of λ of modulus < 1. Then for every $x \in \mathbb{R}^d$*

$$h_{\lambda,\nu_0} \geqslant H(X_{A,x}; A|\mathrm{Id}) \geqslant H(\xi_0 \cdot x; A|\mathrm{Id}),$$
$$h_{\lambda,\nu_0} \geqslant \Phi_{\nu_0}(M_\lambda).$$

We recall that $X_{A,x} := X_A \cdot x$ for $x \in \mathbb{R}^d$ and X_A was defined in (2.1).

Proof. Note that the spectral radius of A is less than 1 and thus the random variable X_A is well defined and bounded. Recall that we denote by $\mu_{A,x}$ and $\mu_{A,x}^{(\ell)}$ the laws of the random variables $X_{A,x}$ and $X_{A,x}^{(\ell)}$ respectively. In what follows it will be convenient for us to write $H(\mu; B)$ for $H(X; B)$, where X is a random variable with law μ. In a similar fashion we also use the notation $H(\mu; B_1|B_2)$.

First we observe that

$$H(\mu_{A,x}; A^\ell|\mathrm{Id}) \leqslant H(\mu_{A,x}^{(\ell)}).$$

Indeed

$$
\begin{aligned}
H(\mu_{A,x}; A^\ell|\mathrm{Id}) &= H(\mu_{A,x}; A^\ell) - H(\mu_{A,x}; \mathrm{Id}) \\
&= H(\mu_{A,x}^{(\ell)} * A^\ell \mu_{A,x}; A^\ell) - H(\mu_{A,x}; \mathrm{Id}) \\
&\leqslant H(\mu_{A,x}^{(\ell)}; A^\ell) + H(A^\ell \mu_{A,x}; A^\ell) - H(\mu_{A,x}; \mathrm{Id}) \\
&\leqslant H(\mu_{A,x}^{(\ell)}; A^\ell) \\
&\leqslant H(\mu_{A,x}^{(\ell)}),
\end{aligned}
$$

where we used the self-similarity relation (2.3) on the second line, the sub-additivity property (ii) in Lemma 9 on the third line. In the fourth line, we used the fact that the gaussian G_A has the same law as AG_{Id} together with the change of variable formula for the entropy (2.8). Finally the last line follows from the sub-additivity property (2.10).

On the other hand, by definition

$$H(\mu_{A,x}; A^\ell|\mathrm{Id}) = \sum_{i=1}^{\ell} H(\mu_{A,x}; A^i|A^{i-1}).$$

We claim that each term in this sum is bounded below by $H(\mu_{A,x}; A|\mathrm{Id})$. This claim follows easily from the self-similarity relation (2.3) and estimate (iv) of Lemma 9. Indeed, our assumption that $||A|| \leqslant 1$ implies that $||A^i y|| \leqslant ||A^{i-1} y||$ for every $y \in \mathbb{R}^d$ and $i \geqslant 1$, and thus ensures that $A^i \preceq A^{i-1}$. This makes the use of Lemma 9 (iv) legitimate and shows that

$$
\begin{aligned}
H(\mu_{A,x}; A^i|A^{i-1}) &= H(\mu_{A,x}^{(i-1)} * A^{i-1}\mu_{A,x}; A^i|A^{i-1}) \\
&\geqslant H(A^{i-1}\mu_{A,x}; A^i|A^{i-1}) = H(\mu_{A,x}; A|\mathrm{Id}).
\end{aligned}
$$

Now to conclude the proof of the first inequality, observe that

$$
\begin{aligned}
h_{\lambda,\nu_0} = \lim_{\ell \to \infty} \frac{1}{\ell} H(\mu_\lambda^{(\ell)}) &\geqslant \lim_{\ell \to \infty} \frac{1}{\ell} H(\mu_{A,x}^{(\ell)}) \\
&\geqslant \lim_{\ell \to \infty} \frac{1}{\ell} H(\mu_{A,x}; A^\ell|\mathrm{Id}) \geqslant H(\mu_{A,x}; A|\mathrm{Id}).
\end{aligned}
$$

To deduce the second bound, we use again Lemma 9 (iv):

$$H(\mu_{A,x}; A|\mathrm{Id}) \geqslant H(\mu_{A,x}^{(1)}; A|\mathrm{Id}) = H(\xi_0 \cdot x; A|\mathrm{Id}).$$

We turn to the proof of the third bound. To this end, we optimize the parameters $x \in \mathbb{R}^d$ and $A \in M_d(\mathbb{R})$. Note that A is allowed to vary among all matrices in $M_d(\mathbb{R})$ with prescribed spectrum: $\{\lambda_1, \ldots, \lambda_d\}$, the Galois conjugates of λ that lie inside the open unit disc, and such that $||A|| \leqslant 1$.

Exploiting the rotational symmetry of the normalized gaussian law, we observe that for any two orthogonal matrices $u, v \in O_d(\mathbb{R})$,

$$H(\xi_0 \cdot x; A | \mathrm{Id}) = H(\xi_0 \cdot ux; uAv | \mathrm{Id}).$$

By the Cartan decomposition, every matrix $A \in M_d(\mathbb{R})$ can be written as $A = uDv$ for a diagonal matrix $D = \mathrm{diag}(\alpha_1, \ldots, \alpha_d)$ with $\alpha_1 \geqslant \ldots \geqslant \alpha_d \geqslant 0$ and $u, v \in O_d(\mathbb{R})$. The α_i's are called the singular values of A. A well-known theorem of Horn ([28] and [29, Pb. 2, p 222]) describes the set of values that can arise as singular values of a matrix with prescribed spectrum. It follows from this result that one can find a real matrix $A \in M_d(\mathbb{R})$ whose eigenvalues are as above $\lambda_1, \ldots, \lambda_d$ and whose singular values are $\alpha_1 = \ldots = \alpha_{d-1} = 1$ and $\alpha_d = |\lambda_1 \cdots \lambda_d| = 1/M_\lambda$. Note that $\alpha_1 = \|A\| = 1$, so this matrix A satisfies our requirements.

In our case we can also find A by the following simple alternative argument: given $v \in \mathbb{R}^d$, consider the quadratic form $q_v(x) := \sum_{n \geqslant 0} \langle A_0^n x, v \rangle^2$, where $\langle \cdot \rangle$ is the canonical Euclidean scalar product on $\mathbb{R}^d$, and A_0 is any diagonalizable matrix with the prescribed eigenvalues. This form is well-defined, because the spectral radius of A_0 is less than 1. Pick $v \in \mathbb{R}^d$ so that the vectors ${}^t A_0^n v$ span $\mathbb{R}^d$ as n varies among the integers (this is always possible since A_0 is invertible and has distinct eigenvalues). Then q_v is positive definite and there exists $B \in \mathrm{GL}_d(\mathbb{R})$ such that $q_v(x) = \|Bx\|^2$. Now by construction:

$$q_v(x) = \langle x, v \rangle^2 + q_v(A_0 x) = \langle x, v \rangle^2 + \|BA_0 x\|^2,$$

from which it follows that $\|BA_0 B^{-1} x\|^2 = \|x\|^2 - \langle x, {}^t B^{-1} v \rangle^2$, which means that at least $d - 1$ singular values of $BA_0 B^{-1}$ are equal to 1. The last one is unambiguously determined by the determinant, hence equal to $1/M_\lambda$.

We write $D = \mathrm{diag}(1, \ldots, 1, M_\lambda^{-1})$. By the above discussion, there are two orthogonal matrices $u, v \in O(d)$ such that $A = uDv$ and hence $D = u^{-1} A v^{-1}$. We deduce $h_{\lambda, \nu_0} \geqslant H(\xi_0 \cdot x; D | \mathrm{Id})$ for all $x \in \mathbb{R}^d$.

We now pick x of the form $x = t e_d$, for some $t \in \mathbb{R}$ and where $e_d = (0, \ldots, 0, 1)$ is the last element of the canonical basis of $\mathbb{R}^d$. Then

$$
\begin{aligned}
H(\xi_0 x; D | \mathrm{Id}) =& H(\xi_0 t e_d + D G_{\mathrm{Id}}) - H(D G_{\mathrm{Id}}) \\
& - (H(\xi_0 t e_d + G_{\mathrm{Id}}) - H(G_{\mathrm{Id}})) \\
=& H(\xi_0 t M_\lambda e_d + G_{\mathrm{Id}}) - H(\xi_0 t e_d + G_{\mathrm{Id}}) \\
=& H(\xi_0 t M_\lambda + G) - H(\xi_0 t + G)
\end{aligned}
$$

where in the last line G is a normalized one-dimensional gaussian random variable. We used the change of variable formula (2.8) to prove the second equality and the identity $F(xy) = xF(y) + yF(x)$ satisfied by the function $F(x) = -x \log(x)$ to integrate out the first $d - 1$ variables. This completes the proof of the proposition. □

3.2. **Proof of the lower bound in Theorem 5.** In this paragraph, we complete the proof of the lower bound in Theorem 5 and to this aim, we study the function

$$\Phi_{\nu_0}(a) = \sup_{t > 0} \{ H(\xi_0 t a + G) - H(\xi_0 t + G) \}$$

and prove the following estimates:

Lemma 14. *The function Φ_{ν_0} is monotone increasing and we have*

$$\frac{\Phi_{\nu_0}(a)}{\log a} \geq \frac{\Phi_{\nu_0}(a^2)}{\log a^2}$$

for any $a > 1$.

This has the following immediate corollary.

Corollary 15. *We have*

$$\Phi_{\nu_0}(a) \geq c \min\{\log a, 1\}$$

for all $a > 1$, where

$$c = \min_{\sqrt{2} \leq a \leq 2} \{\Phi_{\nu_0}(a)/\log(a)\}. \tag{3.1}$$

The constant c can be numerically estimated by calculating $\Phi_{\nu_0}(a)$ for a running through a sufficiently dense arithmetic progression and using the monotonicity of Φ_{ν_0} to estimate it in the intervals between the points of the progression. Our numerical calculations show $c \geq 0.44$ for the case $\nu_0 = (\delta_{-1} + \delta_1)/2$ that is relevant for Bernoulli convolutions. In these calculations, we estimated $\Phi_{\nu_0}(a)$ for a running between 1.4 and 2 in increments of 0.01. We used the lower bound

$$\Phi_{\nu_0}(a) \geq H(\xi_0 t a^{1/2} + G) - H(\xi_0 t a^{-1/2} + G)$$

setting $t = 1.19$. This value was selected to optimize the lower bound at $a = 2$. The calculations used MATLAB's built-in routines for numerical evaluation of integrals, which do not provide error estimates, therefore these calculations are not rigorous. According to our calculations the function $\Phi_{\nu_0}(a)/\log(a)$ appears to be monotone decreasing, hence the minimum is probably attained for $a = 2$.

We note that for each $\varepsilon > 0$, it is possible to choose the measure ν_0 in such a way that the constant (3.1) is at least $1 - \varepsilon$. Indeed, a simple calculation shows that $\Phi_\nu(a) = \log a$, for each $a > 1$ if ν is a Gaussian measure. Hence, one can take for ν_0 a measure supported on the rationals that suitably approximates a Gaussian measure and find that c is as close to 1 as desired.

Observe that Corollary 15 combined with Proposition 13 completes the proof of the lower bound in Theorem 5.

Proof of Lemma 14. Using the change of variable formula (2.8) we get

$$\begin{aligned} H(\xi_0 t a + G) - H(\xi_0 t + G) =& H(\xi_0 t + a^{-1}G) - H(a^{-1}G) \\ &- [H(\xi_0 t + G) - H(G)] \\ =& H(\xi_0 t; a^{-1}|1). \end{aligned}$$

This is an increasing function of a for each fixed t by part (iii) of Lemma 9. This shows that $\Phi_{\nu_0}(a)$ is an increasing function of a.

We turn to the second claim. Fix a and $\varepsilon > 0$ and let t be such that

$$\Phi_{\nu_0}(a^2) \leq H(\xi_0 t a^2 + G) - H(\xi_0 t + G) + \varepsilon.$$

Then by the definition of $\Phi_{\nu_0}(a)$, we have

$$\Phi_{\nu_0}(a^2) \leq H(\xi_0 ta^2 + G) - H(\xi_0 ta + G) + H(\xi_0 ta + G) - H(\xi_0 t + G) + \varepsilon \leq 2\Phi_{\nu_0}(a) + \varepsilon.$$

We take $\varepsilon \to 0$ to conclude $\Phi_{\nu_0}(a^2) \leq 2\Phi_{\nu_0}(a)$, which is precisely the second claim. $\square$

Proof of Corollary 15. If $a \geq 2$, then

$$\Phi_{\nu_0}(a) \geq \Phi_{\nu_0}(2) \geq c = c \min\{\log a, 1\}$$

by the monotonicity of Φ_{ν_0} and the definition of c.

If $a < 2$, then we take $n \geq 0$ to be an integer such that $\sqrt{2} \leq a^{2^n} < 2$. Then by the second claim of Lemma 14 applied repeatedly and the definition of c we have

$$\frac{\Phi_{\nu_0}(a)}{\log a} \geq \frac{\Phi_{\nu_0}(a^{2^n})}{\log a^{2^n}} \geq c.$$

This shows $\Phi_{\nu_0}(a) \geq c \log a$ proving the claim. $\square$

3.3. **Proof of the upper bounds.** The goal of this paragraph is to prove the upper bound $h_\lambda \leq \rho_\lambda \leq \min\{1, \log M_\lambda\}$ in Theorem 5. We will also show Proposition 7, which says that the inequality $h_\lambda \leq \min\{1, \log M_\lambda\}$ is strict if $M_\lambda < 2$ and λ has no Galois conjugate on the unit circle. For simplicity, we assume as we may that ν_0 is supported on $\mathbb{Z}$.

We first recall a simple counting lemma.

Lemma 16. *Let λ be an algebraic unit and denote by k the number of Galois conjugates of λ on the unit circle. Then $|\operatorname{Supp}(\mu_\lambda^{(\ell)})| \leq C\ell^k M_\lambda^\ell$, where C is a constant depending only on λ and ν_0. In particular*

$$\rho_\lambda \leq \min\{1, \log M_\lambda\}.$$

Recall that ρ_λ is defined in (1.4). This lemma is standard, but we give the proof for the reader's convenience.

Proof. We denote by $\sigma_1, \ldots \sigma_n : \mathbb{Q}(\lambda) \to \mathbb{R}$ the real Galois embeddings with $|\sigma_i(\lambda)| \geq 1$ and by $\tau_1, \ldots, \tau_m : \mathbb{Q}(\lambda) \to \mathbb{C}$ the complex Galois embeddings with $|\tau_i(\lambda)| \geq 1$ such that we take exactly one from each pair of complex conjugate embeddings. Furthermore, we denote by $\rho_1, \ldots, \rho_o : \mathbb{Q}(\lambda) \to \mathbb{C}$ the real or complex Galois embeddings with $|\rho_i(\lambda)| < 1$. (Here we take both from a pair of complex conjugate embeddings.) We define the map $S : \mathbb{Q}(\lambda) \to \mathbb{R}^{n+2m}$ by

$$S(x) = (\sigma_1(x), \ldots, \sigma_n(x), \operatorname{Re}(\tau_1(x)), \operatorname{Im}(\tau_1(x)), \ldots, \operatorname{Re}(\tau_m(x)), \operatorname{Im}(\tau_m(x))).$$

We consider the set

$$A = \operatorname{Supp}(\mu_\lambda^{(\ell)}) = \left\{ \sum_{i=0}^{\ell-1} a_i \lambda^i : a_i \in \operatorname{Supp}\nu_0; \text{for all } i \right\}.$$

We note that elements of A are algebraic integers, hence for any two different $x, y \in A$ we have

$$\prod_{i,j,k} |\sigma_i(x-y)||\tau_j(x-y)|^2|\rho_k(x-y)| \geq 1.$$

For any $1 \leq k \leq o$, we have $|\rho_k(x-y)| \leq |\rho_k(x)| + |\rho_k(y)| \leq 2L/(1-|\rho_k(\lambda)|)$, where $L = \max_{x \in \text{Supp}\, \nu_0} |x|$. Hence there is a number c depending on λ and L such that

$$\prod_{i,j} |\sigma_i(x-y)||\tau_j(x-y)|^2 \geq c.$$

Thus $\|S(x-y)\| > c_1$ for some other number c_1 depending on λ and L.

Consider the set

$$\Omega = \{(x_1, \ldots, x_{n+2m}) \in \mathbb{R}^{n+2m} :$$

$$|x_i| \leq L \frac{|\sigma_i(\lambda)|^\ell - 1}{|\sigma_i(\lambda)| - 1} + c_1 \text{ for } 1 \leq i \leq n, \text{ while for } 1 \leq j \leq m,$$

$$|x_{n+2j-1}|, |x_{n+2j}| \leq L \frac{|\tau_j(\lambda)|^\ell - 1}{|\tau_j(\lambda)| - 1} + c_1 \text{ if } |\tau_j(\lambda)| > 1,$$

$$|x_{n+2j-1}|, |x_{n+2j}| \leq L\ell + c_1 \text{ if } |\tau_j(\lambda)| = 1\}.$$

It is easily seen that the balls of radii $c_1/2$ around the points $S(x)$ for $x \in A$ are disjoint and contained in Ω. On the other hand, there is a number C depending only on λ and L such that the volume of Ω is less than $C\ell^k M_\lambda^\ell$, hence the claim follows. $\square$

Proof of the upper bound in Theorem 5. Using (2.9) we see that $h_\lambda \leqslant \rho_\lambda$, so the bound follows from Lemma 16. $\square$

It remains to prove Proposition 7, which we recall now:

Proposition 17. *Let λ be an algebraic number such that $M_\lambda < 2$ and assume that λ has no conjugates on the unit circle. Then $h_\lambda < \log M_\lambda$.*

From now on we restrict ourselves to the special case $\nu_0 = (\delta_{-1} + \delta_1)/2$ that is relevant to Bernoulli convolutions. The proof is based on ideas from Garsia's proof that $h_\lambda < -\log \lambda$ if λ^{-1} is Pisot [23].

We use the following notation (note that it differs from that of the proof of Lemma 16). We denote by $\sigma_1, \ldots \sigma_n : \mathbb{Q}(\lambda) \to \mathbb{R}$ the real Galois embeddings with $|\sigma_i(\lambda)| < 1$ and by $\tau_1, \ldots, \tau_m : \mathbb{Q}(\lambda) \to \mathbb{C}$ the complex Galois embeddings with $|\tau_i(\lambda)| < 1$ such that we take exactly one from each pair of complex conjugate embeddings. Furthermore, we denote by $\rho_1, \ldots, \rho_o : \mathbb{Q}(\lambda) \to \mathbb{C}$ the real or complex Galois embeddings with $|\rho_i(\lambda)| > 1$. (Here we take both from a pair of complex conjugate embeddings.) We define the map $S : \mathbb{Q}(\lambda) \to \mathbb{R}^{n+2m}$ by

$$S(x) = (\sigma_1(x), \ldots, \sigma_n(x), \text{Re}(\tau_1(x)), \text{Im}(\tau_1(x)), \ldots, \text{Re}(\tau_m(x)), \text{Im}(\tau_m(x))).$$

We introduce the random vectors

$$Y_\lambda = \sum_{i=0}^{\infty} \xi_i S(\lambda^i) = S(X_\lambda), \quad Y_\lambda^{(\ell)} = \sum_{i=0}^{\ell-1} \xi_i S(\lambda^i) = S(X_\lambda^{(\ell)}).$$

The strategy of the proof of the proposition is the following. We begin by proving that the law of Y_λ is singular (Lemma 18). Then we approximate the law of $Y_\lambda^{(\ell)}$ by Y_λ and conclude that most of the probability mass is concentrated on an ε proportion of the atoms. This yields a slight improvement over the proof of the upper bound in Theorem 5, which is just enough to conclude that $h_\lambda < M_\lambda$.

Lemma 18. *Suppose that λ is an algebraic unit that has no Galois conjugates on the unit circle. Then the law of Y_λ is singular.*

The proof is a straightforward generalization of the fact that μ_λ is singular if λ^{-1} is Pisot.

Proof. We put for each $t \in \mathbb{N}$

$$\zeta_t = (\sigma_1(\lambda^{-t}), \ldots, \sigma_n(\lambda^{-t}), 2\operatorname{Re}(\tau_1(\lambda^{-t})), -2\operatorname{Im}(\tau_1(\lambda^{-t})), \ldots,$$
$$2\operatorname{Re}(\tau_m(\lambda^{-t})), -2\operatorname{Im}(\tau_m(\lambda^{-t}))).$$

We show below that there is a number $c > 0$ depending only on λ such that

$$|\mathbb{E}[\exp(2\pi i \langle \zeta_t, Y_\lambda \rangle)]| \geq c.$$

Then the Riemann-Lebesgue Lemma implies that the law of Y_λ is not absolutely continuous. Since the law is self-affine, it is of pure type, hence it is singular.

We observe that

$$\mathbb{E}[\exp(2\pi i \langle \zeta_t, Y_\lambda \rangle)] = \prod_{j=0}^{\infty} \mathbb{E}[\exp(2\pi i \langle \zeta_t, \xi_j S(\lambda^j) \rangle)]$$
$$= \prod_{j=0}^{\infty} \cos(2\pi \langle \zeta_t, S(\lambda^j) \rangle).$$

We can write

$$\langle \zeta_t, S(\lambda^j) \rangle = \sum_{a=1}^{n} \sigma_a(\lambda^{-t}) \sigma_a(\lambda^j)$$
$$+ 2 \sum_{b=1}^{m} (\operatorname{Re}(\tau_b(\lambda^{-t})) \operatorname{Re}(\tau_b(\lambda^j)) - \operatorname{Im}(\tau_b(\lambda^{-t})) \operatorname{Im}(\tau_b(\lambda^j)))$$
$$= \sum_{a=1}^{n} \sigma_a(\lambda^{j-t}) + 2 \sum_{b=1}^{m} \operatorname{Re}(\tau_b(\lambda^{j-t})).$$

We set

$$u_j = \sum_{a=1}^{n} \sigma_a(\lambda^j) + 2 \sum_{b=1}^{m} \operatorname{Re}(\tau_b(\lambda^j)),$$
$$v_j = \sum_{a=1}^{o} \rho_a(\lambda^j).$$

Since λ is a unit and $u_j + v_j$ equals to the sum of all Galois conjugates of λ^j, it follows that $u_j + v_j$ is an integer for all $j \in \mathbb{Z}$. Hence we can write

$$|\mathbb{E}[\exp(2\pi i \langle \zeta_t, Y_\lambda \rangle)]| \geq \prod_{j=-\infty}^{\infty} |\cos(2\pi u_j)|$$
$$= \prod_{j=-\infty}^{-1} |\cos(2\pi v_j)| \prod_{j=0}^{\infty} |\cos(2\pi u_j)| =: c.$$

We now show that the quantity of the right hand side, which we have denoted by c is strictly positive. We note that there is a positive number

$\alpha < 1$ depending only on λ such that $|u_j| \leq (n+2m)\alpha^j$ and $|v_j| \leq o\alpha^{-j}$ for all $j \in \mathbb{Z}$. We choose an integer N large enough so that $(n+2m)\alpha^j < 1/4$ and $o\alpha^j < 1/4$ for $j > N$. There is an absolute constant $C > 0$ such that $\cos(x) \geq \exp(-C|x|)$ for $|x| < 1/4$, hence we can write

$$c \geq \exp\Big(-Co\sum_{j=-\infty}^{-N-1} \alpha^{-j}\Big) \prod_{j=-N}^{N} |\cos(2\pi u_j)| \exp\Big(-C(n+2m)\sum_{j=N+1}^{\infty} \alpha^j\Big).$$

We note that u_j is an algebraic integer for all $j \in \mathbb{Z}$ so it cannot be equal to one half plus an integer. This finishes the proof that $c > 0$, hence the lemma follows. $\square$

Proof of Proposition 17. Denote by Q_ℓ the box in $\mathbb{R}^{n+2m}$ with side lengths

$$|\sigma_1(\lambda)|^\ell, \ldots, |\sigma_n(\lambda)|^\ell, |\tau_1(\lambda)|^\ell, |\tau_1(\lambda)|^\ell, |\tau_2(\lambda)|^\ell, |\tau_2(\lambda)|^\ell \ldots, |\tau_m(\lambda)|^\ell, |\tau_m(\lambda)|^\ell$$

centered around the origin.

We begin by an observation about the separation of points in the support of $Y_\lambda^{(\ell)}$. Let $x, y \in \operatorname{Supp} X_\lambda^{(\ell)}$ be two different points. We can apply an argument similar to that in the proof of Lemma 16 for the points $S(\lambda^{-(\ell-1)}x)$, $S(\lambda^{-(\ell-1)}y)$ and λ^{-1} in place of λ to show that there is a number $c > 0$ depending only on λ such that $S(\lambda^{-(\ell-1)}x) - S(\lambda^{-(\ell-1)}y) \notin cQ_0$. By cQ_0 and by similar notation, we mean the dilation or contraction of Q_0 by the factor c. This in turn yields $S(x) - S(y) \notin cQ_{\ell-1}$.

Next, we estimate the difference between Y_λ and $Y_\lambda^{(\ell)}$. We can write

$$|\sigma_i(X_\lambda - X_\lambda^{(\ell)})| = \Big|\sum_{j=\ell}^{\infty} \sigma_i(\lambda^j)\Big| \leq C|\sigma_i(\lambda)|^\ell,$$

where C is a constant depending only on λ. A similar inequality holds for the embeddings τ_i. We can conclude hence that $Y_\lambda - Y_\lambda^{(\ell)} \in CQ_\ell$.

We fix a small number $\varepsilon > 0$. Since the law of Y_λ is singular, we can find a closed set $A \subset \mathbb{R}^{n+2m}$ such that $\mu(A) < \varepsilon$ and $\mathbb{P}(Y_\lambda \in A) > 1 - \varepsilon$. Here and everywhere below, μ denotes the Lebesgue measure.

If ℓ is sufficiently large, we have $\mu(A + (C + \frac{c}{2})Q_\ell) < 2\varepsilon$, where c and C are the same as above, because A is closed. We estimate the cardinality of $A' := \operatorname{Supp}(Y_\lambda^{(\ell)}) \cap (A + CQ_\ell)$. If $x, y \in A'$ are distinct, then $x + (c/2)Q_\ell$ and $y + (c/2)Q_\ell$ are disjoint. Hence

$$|A'| = \frac{\mu(A' + (c/2)Q_\ell)}{\mu((c/2)Q_\ell)} < C_1 \varepsilon M_\lambda^\ell,$$

where C_1 is a constant depending only on λ. By Lemma 16 (here $k = 0$), we can estimate from above the cardinality of $B' := \operatorname{Supp}(Y_\lambda^{(\ell)}) \setminus A'$ by $C_2 M_\lambda^\ell$. Moreover $\mathbb{P}(Y_\lambda^{(\ell)} \in B') \leq \mathbb{P}(Y_\lambda \notin A) < \varepsilon$.

We put our estimates together to bound $H(Y_\lambda^{(\ell)})$. We write

$$
\begin{aligned}
H(Y_\lambda^{(\ell)}) = \sum_{x \in A'} &-\mathbb{P}(Y_\lambda^{(\ell)} = x) \log(\mathbb{P}(Y_\lambda^{(\ell)} = x)) \\
&+ \sum_{x \in B'} -\mathbb{P}(Y_\lambda^{(\ell)} = x) \log(\mathbb{P}(Y_\lambda^{(\ell)} = x)) \\
\leq |A'| \cdot &\Big(-\frac{\mathbb{P}(Y_\lambda^{(\ell)} \in A')}{|A'|} \log \Big(\frac{\mathbb{P}(Y_\lambda^{(\ell)} \in A')}{|A'|} \Big) \Big) \\
&+ |B'| \cdot \Big(-\frac{\mathbb{P}(Y_\lambda^{(\ell)} \in B')}{|B'|} \log \Big(\frac{\mathbb{P}(Y_\lambda^{(\ell)} \in B')}{|B'|} \Big) \Big) \\
= -&\mathbb{P}(Y_\lambda^{(\ell)} \in A') \log(\mathbb{P}(Y_\lambda^{(\ell)} \in A')) - \mathbb{P}(Y_\lambda^{(\ell)} \in B') \log(\mathbb{P}(Y_\lambda^{(\ell)} \in B')) \\
&+ \mathbb{P}(Y_\lambda^{(\ell)} \in A') \log|A'| + \mathbb{P}(Y_\lambda^{(\ell)} \in B') \log|B'| \\
\leq 1 + &(1-p)\log(C_1 \varepsilon M_\lambda^\ell) + p\log(C_2 M_\lambda^\ell) = \log(2C_1 \varepsilon M_\lambda^\ell) - p\log(\frac{C_1}{C_2}\varepsilon),
\end{aligned}
$$

where $p = \mathbb{P}(Y_\lambda^{(\ell)} \in B') \leqslant \varepsilon$. The inequality in the second line follows from the concavity of the function $F(x) = -x\log(x)$. If we set ε sufficiently small depending only on C_1 and C_2, hence ultimately depending only on λ, then we obtain $H(Y_\lambda^{(\ell)}) < \log(M_\lambda^\ell)$. We can conclude now that

$$
h_\lambda \leq \frac{H(Y_\lambda^{(\ell)})}{\ell} < \log(M_\lambda)
$$

proving the claim. $\square$

3.4.   Although (1.2) is not stated in this form in [27] it is essentially contained in that paper. For the reader's convenience, we show how to reduce it to the main result of [27] that we now recall. All of the ideas in this paragraph are taken from [27].

For an integer n denote by $\mathcal{D}_n$ the partition of $\mathbb{R}$ into intervals of length 2^{-n} such that 0 is an endpoint of two intervals in the partition and by $H(\nu, \mathcal{D}_n)$ the Shannon entropy of ν with respect to the partition $\mathcal{D}_n$. For integers $m < n$, denote by $H(\nu, \mathcal{D}_n | \mathcal{D}_m) = H(\nu, \mathcal{D}_n) - H(\nu, \mathcal{D}_m)$ the conditional entropies. This is the notation in [27], which differs from ours, and we only use it in this paragraph.

Theorem 19 (Special case of [27, Theorem 1.3]). *Let $\lambda \in (1/2, 1)$ and suppose that $\dim \mu_\lambda < 1$. Then*

$$
\lim_{n \to \infty} \frac{1}{n'} H(\mu_\lambda^{(n)}, \mathcal{D}_{qn'} | \mathcal{D}_{n'}) = 0 \tag{3.2}
$$

for any $q > 0$, where $n' = \lfloor n \cdot \log \lambda^{-1} \rfloor$.

We first consider the case when $\dim \mu_\lambda < 1$. We recall from [22, Lemma 1.52] that for each algebraic number λ, there are c_λ and d_λ such that any two distinct points in the support of $\mu_\lambda^{(n)}$ are of distance at least $c_\lambda n^{-d_\lambda} M_\lambda^{-n}$. Hence, taking q sufficiently large, we can write

$$
\lim_{n \to \infty} \frac{1}{n} H(\mu_\lambda^{(n)}, \mathcal{D}_{qn'}) = \lim_{n \to \infty} \frac{1}{n} H(\mu_\lambda^{(n)}) = h_\lambda.
$$

We combine this with (3.2) and deduce

$$\lim_{n\to\infty} \frac{1}{n'} H(\mu_\lambda^{(n)}, \mathcal{D}_{n'}) = \frac{h_\lambda}{\log \lambda^{-1}}.$$

Since $\mu_\lambda^{(n)}$ approximates μ_λ at scale $2^{-n'}$,

$$\lim_{n\to\infty} \frac{1}{n'} H(\mu_\lambda, \mathcal{D}_{n'}) = \lim_{n\to\infty} \frac{1}{n'} H(\mu_\lambda^{(n)}, \mathcal{D}_{n'})$$

The quantity on the left hand side of this equation is known to equal $\dim \mu_\lambda$, see [21].

It is left to consider the case $\dim \mu_\lambda = 1$, and we need to prove that $h_\lambda \geq \log \lambda^{-1}$. The observation that $h_\lambda < \log \lambda^{-1}$ implies μ_λ is singular goes back to Garsia [23]. Here we need a slightly stronger statement that we obtain by writing

$$1 = \dim \mu_\lambda = \lim_{n\to\infty} \frac{1}{n'} H(\mu_\lambda, \mathcal{D}_{n'}) = \lim_{n\to\infty} \frac{1}{n'} H(\mu_\lambda^{(n)}, \mathcal{D}_{n'})$$
$$\leq \lim_{n\to\infty} \frac{1}{n'} H(\mu_\lambda^{(n)}) = \frac{h_\lambda}{\log \lambda^{-1}}.$$

This completes the proof of (1.2).

We can now prove Theorem 1.

Proof of Theorem 1. If Lehmer's conjecture holds, then there is a number $\varepsilon > 0$ such that $\log M_\lambda > \varepsilon$ for all algebraic numbers $\lambda \in (1/2, 1)$. Then $h_\lambda \geq 0.44\varepsilon$ by Theorem 5. If λ is sufficiently close to 1 so that $\log \lambda^{-1} < 0.44\varepsilon$, then $h_\lambda > \log \lambda^{-1}$ and $\dim \mu_\lambda = 1$ by (1.2). $\qquad \square$

4. UNIFORM EXPONENTIAL GROWTH FOR LINEAR GROUPS

This section is devoted to the proof of Theorem 4 and the consequences of Theorem 5 for group growth. Recall that given a group G generated by a finite subset S, we denote by ρ_S the rate of exponential growth:

$$\rho_S := \lim_{n\to+\infty} \frac{1}{n} \log |S^n|.$$

Since $|S^{n+m}| \leqslant |S^n||S^m|$ for every $n, m \in \mathbb{N}$ the above limit exists, by the classical subadditive lemma. We also note that $\rho_S = \rho_{S\cup\{1\}}$, because $(S \cup \{1\})^n = \bigcup_{i=0}^n S^i$, and hence $|(S \cup \{1\})^n| \leqslant (n+1)|S^n|$. So without loss of generality, we will assume that $1 \in S$.

Before going into any details we record here the following initial observation, whose proof we leave to the reader. Let $g_{\lambda,1}$ and $g_{\lambda,-1}$ be the affine transformations of the complex line $x \mapsto \lambda x + 1$ and $x \mapsto \lambda x - 1$. Set $S_\lambda := \{g_{\lambda,1}, g_{\lambda,-1}\}$. Then we have:

$$\rho_{S_\lambda} = \rho_\lambda,$$

where ρ_λ is defined in (1.4). In particular Theorem 5 combined with (1.5) implies that $\rho_{S_\lambda} \geqslant 0.44 \min\{1, \log M_\lambda\}$.

Recall that an abstract group is said to satisfy a certain property $\mathcal{P}$ *virtually*, or equivalently to be *virtually* $\mathcal{P}$, if it contains a subgroup of finite index with the said property $\mathcal{P}$. For example a group is virtually trivial if and only if it is finite.

Recall further that a group G is said to be *solvable* if the derived series of the group stabilises to the trivial group in finitely many steps, namely setting $G_1 = G$ and recursively $G_{n+1} := [G_n, G_n]$ the subgroup generated by all commutators $aba^{-1}b^{-1}$, $a, b \in G_n$ there is $n < \infty$ such that $G_n = \{1\}$. Similarly a group is said to be *nilpotent* if the central descending series stabilises to the trivial group in finitely many steps, namely setting $G^{(1)} = G$ and recursively $G^{(n+1)} := [G, G^{(n)}]$ the subgroup generated by all commutators $aba^{-1}b^{-1}$, $a \in G$ and $b \in G_n$ there is $n < \infty$ such that $G^{(n)} = \{1\}$. Examples of solvable groups include the group $\mathrm{Upp}_d(\mathbb{C})$ of upper triangular invertible matrices of size d. Examples of nilpotent groups include the commutator subgroup of $\mathrm{Upp}_d(\mathbb{C})$, i.e. the upper triangular and unipotent matrices (i.e. matrices all of whose eigenvalues are 1).

According to a celebrated lemma of Jordan [32], there is a function $J = J(d) \in \mathbb{N}$ such that every finite subgroup of $\mathrm{GL}_d(\mathbb{C})$ contains a normal abelian subgroup of index at most $J(d)$. For this and for general background on linear groups we refer the reader to standard books [17, 43, 49].

The following is the main theorem of this section:

Theorem 20. *Let S be a finite subset of $\mathrm{GL}_d(\mathbb{C})$ generating a virtually solvable subgroup, then either $\rho_S = 0$ and $\langle S \rangle$ is virtually nilpotent, or there is $\lambda \in \mathbb{C}^\times$, not a root of unity, such that*

$$\rho_S \geqslant \frac{1}{27d!J(d)} \log M_\lambda.$$

Recall that M_λ is the Mahler measure of the minimal polynomial of λ in $\mathbb{Z}[X]$ if λ is algebraic over $\mathbb{Q}$. We adopt the convention that $M_\lambda = 2$ if λ is transcendental. We thus obtain the following consequence.

Corollary 21. *Assuming the Lehmer conjecture, there is $c > 0$ such that the following holds. If the finite subset $S \subset \mathrm{GL}_d(\mathbb{C})$ generates a virtually solvable subgroup, then either $\rho_S = 0$ and $\langle S \rangle$ is virtually nilpotent, or*

$$\rho_S > \frac{c}{d!J(d)}.$$

This completes the proof of the equivalence between the Lehmer conjecture and the Growth conjecture (Theorem 4), see the Introduction.

Remark 22. The classical geometric proofs of Jordan's lemma by Bieberbach and Frobenius give a bound on $J(d)$ of the form $d^{O(d/\log d)^2}$, while using the classification of finite simple groups much better bounds have been obtained by B. Weisfeiler and then by M. Collins (see [15]) who shows the sharp bound $J(d) \leqslant (d+1)!$ when $d \geqslant 71$.

Remark 23. It is worth remarking here that the standard argument for proving exponential word growth in finitely generated groups is to exhibit two elements that generate a free semi-group. Every non virtually nilpotent solvable group contains a free semi-group (see e.g. [12] and references therein). However there is no uniform bound on the word length of these free generators: indeed in [12] a construction is given of a sequence of algebraic numbers λ_n such that S_{λ_n} generates a non-virtually nilpotent subgroup of affine transformations, and yet no pair of elements in $(S_{\lambda_n})^n$ generate a free

semi-group. Therefore there is no hope of obtaining a good lower estimate for ρ_S using ping-pong techniques only, as was done in [11, 14] in the non-virtually solvable case. Instead Theorem 20 will be a consequence of the entropy lower bounds established in the first part of this paper.

We will show additionally, that if K_S denotes the field generated by the matrix entries of each element $s \in S$, then λ can be chosen to be algebraic over K_S of degree at most $d!$. Using Dobrowolski's bound for the Mahler measure of an algebraic number of bounded degree (see [18]) we obtain in a similar way the following consequence, which was pointed out to us by Andreas Thom.

Corollary 24. *There is an absolute constant $c > 0$ such that if the finite subset $S \subset \mathrm{GL}_d(\mathbb{Q})$ generates a virtually solvable subgroup, then either $\rho_S = 0$ and $\langle S \rangle$ is virtually nilpotent, or*

$$\rho_S > \frac{c}{d^{2d}}.$$

In the proof of Theorem 20, we first show that Γ contains a finite index subgroup H that can be conjugated into $\mathrm{Upp}_d(\mathbb{C})$, the subgroup of $\mathrm{GL}_d(\mathbb{C})$ made of upper triangular matrices. Then we show that H has a non-virtually solvable image under a suitable homomorphism into $\mathrm{Aff}(\mathbb{C})$, the group of affine transformations of the complex line (similarities). This allows us to reduce the theorem to the special case of $S \subset \mathrm{Aff}(\mathbb{C})$, which we treat first.

Note that we have an isomorphism:

$$\mathrm{Aff}(\mathbb{C}) \simeq \{g_{a,b} := \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix}; a \in \mathbb{C}^*, b \in \mathbb{C}\} \leqslant \mathrm{GL}_2(\mathbb{C})$$

The matrix $g_{a,b}$ identifies with the affine transformation $x \mapsto ax + b$. If $g = g_{a,b}$, we set $a(g) = a$ and $b(g) = b$.

Lemma 25. *Let S be a finite subset of $\mathrm{Aff}(\mathbb{C})$ containing the identity. If $\langle S \rangle$ is not virtually nilpotent, then there is $g, \gamma \in S^3$ such that the pair (g, γ) is conjugate in $\mathrm{Aff}(\mathbb{C})$ to the pair $(g_{\lambda,1}, g_{\lambda,-1})$ for some $\lambda \in \mathbb{C}^*$, which is not a root of unity.*

Proof. The proof relies on the following fact, whose proof we leave to the reader. If $\lambda \in \mathbb{C} \setminus \{0, 1\}$ and $t_1 \neq t_2$, $u_1 \neq u_2$, then the pair $(g_{\lambda,t_1}, g_{\lambda,t_2})$ is conjugate to the pair $(g_{\lambda,u_1}, g_{\lambda,u_2})$ by an element in $\mathrm{Aff}(\mathbb{C})$.

First observe that the multiplicative subgroup of $\mathbb{C}^*$ generated by the $a(s)$, $s \in S$, is infinite, for otherwise the subgroup $\langle S \rangle$ would be virtually abelian. In particular, there is $s_0 \in S$ such that $\lambda_0 := a(s_0)$ is not a root of unity. Up to conjugating S in $\mathrm{Aff}(\mathbb{C})$, we may assume without loss of generality that $b(s_0) = 0$. So $s_0 = g_{\lambda_0,0}$.

Now note that there must exist some $s \in S$ such that $b(s) \neq 0$, for otherwise $\langle S \rangle$ would be abelian. Now not both $\lambda_0 a(s)$ and $\lambda_0^2 a(s)$ are roots of unity. Let γ be either $s_0 s$ or $s_0^2 s$, so that $a(\gamma)$ is not a root of unity. Accordingly, let g be either $s s_0$, or $s s_0^2$.

Then $\lambda := a(\gamma) = a(g)$ is not a root of unity, while $b(\gamma) \neq b(g)$. From the above fact, we deduce that (g, γ) is conjugate in $\mathrm{Aff}(\mathbb{C})$ to the pair $(g_{\lambda,1}, g_{\lambda,-1})$ as desired. $\square$

If $S \subset \mathrm{GL}_d(\mathbb{C})$, recall that K_S denotes the subfield of $\mathbb{C}$ generated by the matrix entries of each $s \in S$.

Corollary 26. *Let S be a finite subset of* $\mathrm{Aff}(\mathbb{C})$ *containing the identity. Assume that $\langle S \rangle$ is not virtually nilpotent, then there is $\lambda \in K_S \setminus \{0\}$, which is not a root of unity, such that*

$$\rho_S \geqslant \frac{1}{9} \log M_\lambda.$$

Proof. After replacing S by a conjugate, S^{3n} contains all products of at most n affine transformations of the form $g_{\lambda,1}$ or $g_{\lambda,-1}$ for some $\lambda \in K_S$. The images of 0 under these maps form the support of the measure $\mu_\lambda^{(n)}$, so $\rho_S \geqslant \frac{1}{3}\rho_\lambda$. The desired inequality then follows from Theorem 5. If $M_\lambda \geqslant 2$, set $\lambda = 2$. $\square$

In what follows will make use of certain notions from the theory of linear algebraic groups for which we refer to the textbook [30]. For the reader's convenience we briefly review some of the terminology.

A subset of $M_n(\mathbb{C})$ is said to be Zariski-closed if it is the set of zeroes of a family of polynomials in the matrix entries. For example $\mathrm{GL}_n(\mathbb{C})$ is viewed as the Zariski closed subset of $M_{n+1}(\mathbb{C})$ of bloc diagonal matrices $\mathrm{diag}(A, x)$, $A \in \mathrm{GL}_n(\mathbb{C})$, $x \in \mathbb{C}$ such that $\det(A)x = 1$. This endows $\mathrm{GL}_n(\mathbb{C})$ with a non-Hausdorff topology called the Zariski topology.

The Zariski closure of a subset is the smallest (i.e. the intersection of all the) Zariski closed subset containing it. A Zariski closed subset is called irreducible if it is not the union of two proper Zariski closed subsets. Every Zariski closed subset is the union of finitely many irreducible Zariski closed subsets called its irreducible compoments. There is a well-defined notion of dimension of a Zariski-closed subset. Zariski closed subgroups are closed complex Lie subgroups of $\mathrm{GL}_n(\mathbb{C})$ and their Zariski dimension coincides with the complex dimension of their Lie algebras.

The irreducible components of a Zariski closed subgroup $G \leqslant \mathrm{GL}_n(C)$ are disjoint: they are the cosets of the unique irreducible component containing the identity, called the connected component of the identity and denoted by G°. A Zariski closed subgroup G is said to be connected if $G = G^\circ$. The Zariski closure of a subgroup (or sub-semi-group) of $\mathrm{GL}_n(\mathbb{C})$ is a group.

A unipotent subgroup is a subgroup made entirely of unipotent elements. Since we are in characteristic zero, every Zariski closed unipotent subgroup is connected ([30, p. 101]). The union of all Zariski closed unipotent normal subgroups of a Zariski closed subgroup $G \leqslant \mathrm{GL}_n(\mathbb{C})$ is itself a Zariski closed unipotent normal subgroup, called the unipotent radical of G and denoted by G_u. A subgroup $H \leqslant \mathrm{GL}_n(\mathbb{C})$ is said to be diagonalizable if it can be conjugated inside the subgroup of diagonal matrices. A Zariski closed and connected diagonalizable subgroup of G of maximal dimension is called a maximal torus. Any two maximal tori are conjugate in G ([30, §21]).

Having recalled this terminology we can now state a technical result about subgroups of $\mathrm{GL}_n(\mathbb{C})$ that can be conjugated into $\mathrm{Upp}_n(\mathbb{C})$. This will be used both for finding the finite index subgroup in Γ that can be conjugated into $\mathrm{Upp}_n(\mathbb{C})$ and for finding the non-virtually solvable homomorphic image of that group in $\mathrm{Aff}(\mathbb{C})$.

Lemma 27. *Let G be a solvable Zariski closed algebraic subgroup of $\mathrm{GL}_n(\mathbb{C})$, and let G° be the connected component of the identity. The following are equivalent:*

(1) *$[G, G]$ is unipotent,*
(2) *G is a subgroup of a connected solvable algebraic subgroup of $\mathrm{GL}_n(\mathbb{C})$,*
(3) *G can be conjugated into $\mathrm{Upp}_n(\mathbb{C})$.*

Moreover if this holds, then there is a finite abelian subgroup F such that $G = FG^\circ$, and a diagonalizable subgroup $H \leqslant G$ containing F such that $H = FH^\circ$, and $G^\circ = H^\circ G_u$ where G_u is the unipotent radical of G°.

Proof. The equivalence of (2) and (3) is the content of the Lie-Kolchin theorem, see [30, 17.6], and (1) trivially follows from (3).

In order to show that (1) implies (3), we first show that the groups H and F with the properties stated in the lemma exist. First recall that it is a well-known observation attributed to Platonov [42] (see also [49, 10.10], [5, 5.11]) that every complex linear algebraic group has a finite subgroup intersecting each irreducible component. So there is a finite subgroup F with $G = FG^\circ$. Assuming (1) and the existence of F such that $G = FG^\circ$, we will prove the existence of H as above by induction on $\dim G$.

If (1) holds then $[G, G]$ is contained in G_u (its Zariski closure is a closed normal unipotent subgroup of G). This implies that F must be abelian: $[F, F]$ is finite and unipotent, hence trivial. Being finite, F must consist of semisimple (i.e. diagonalizable) elements, and since it is abelian, it is a diagonalizable subgroup.

To find H, we argue as in the standard proof of the existence of a maximal torus mapping onto the quotient of a connected solvable algebraic group with its unipotent radical ([30, 19.3]). Let $Z_G(F)$ be the centralizer of F in G and $Z_G(F)^\circ$ its connected component of the identity. Since $[F, G] \subset G_u$, the diagonalizable subgroup F acts trivially by conjugation on G/G_u, and [30, Corollary 18.4] tells us that the map $\phi : G \to G/G_u$ sends $Z_G(F)^\circ$ onto G°/G_u.

It follows that $G = FG^\circ = FZ_G(F)^\circ G_u$, so $Z_G(F) = FZ_G(F)^\circ(G_u \cap Z_G(F))$. But Zariski closed unipotent subgroups are connected, so $G_u \cap Z_G(F) \leqslant Z_G(F)^\circ$ and we conclude that $Z_G(F) = FZ_G(F)^\circ$. So we may replace G by $Z_G(F)$ and argue by induction if $\dim Z_G(F) < \dim G$. The subgroup H thus found for $Z_G(F)$ will work for G as well. Otherwise $G^\circ \leqslant Z_G(F)$, and given a maximal torus in $H^\circ \leqslant G^\circ$, we may set $H := FH^\circ$, which is the desired diagonalizable subgroup. Since G° is connected and solvable we have $G^\circ = H^\circ G_u$ by [30, §19.3].

Therefore, H and F exist as in the statement of the lemma. Since G_u is unipotent and normal (even characteristic) in G, its fixed point subspace is non-trivial and G-invariant. In fact there is a flag of G-invariant subspaces such that G_u acts trivially on each successive quotient. In particular H preserves this flag and it and can thus be diagonalized in an adapted basis. In that basis $G = HG_u$ is upper triangular, and (3) follows. $\square$

Remark 28. Denote by K the field of definition of the algebraic group G that appears in the previous lemma. It is not a priori clear that the subgroup F in the conclusion of the lemma can be chosen inside $G(K)$. However, we

show now that this is indeed the case if $G \leqslant \mathrm{Upp}_n$, and moreover H can be chosen defined and split over K. This observation is not needed for the proof of the main result Theorem 20, but it will be used for Corollary 24.

To see it, we need to go back to the proof of Platonov's obervation, as given for example in [49, §10]. We first reduce to the case when G° is nilpotent. Note that since maximal K-tori in Upp_n are K-split (i.e. isomorphic to $(\mathbb{C}^{\times})^n$ via an isomorphism defined over K), so are the maximal K-tori T in G°. Since they are all conjugate by an element of $G_u(K)$, we have $G(K) = N_G(T)(K)G_u(K)$. But $N_G(T) = Z_G(T)$ (apply [30, 19.4.b] with $G = \mathrm{Upp}_n$), so replacing G with $Z_G(T)$, we may assume that maximal K-tori in G° are K-split and are inside the center of G. This implies in particular that there is only one such T and $G^{\circ} = TG_u$ is nilpotent.

Then G/T is virtually unipotent, and $G^{\circ}(K)/T(K)$ is divisible, torsion free and nilpotent. Now it follows, as in the discussion [49, §10.10], that the exact sequence $1 \to G_u(K) \to G(K)/T(K) \to G(K)/G^{\circ}(K) \to 1$ splits.

So there is a subgroup $F \subset G(K)$ such that $F \cap G_u(K) = 1$ (forcing F to be abelian) and $G(K) = FT(K)G_u(K)$. Clearly F is made of semisimple elements (if $f \in F$ some power of f lies in T). Also every semisimple element in $\mathrm{Upp}_d(K)$ is diagonalizable. It follows that $FT(K)$ is an abelian subgroup made of diagonalizable elements, hence it can be simultaneously diagonalized, and this yield the desired subgroup H.

We now move towards the proof of Theorem 20. We will use the previous lemma to reduce to the case of the 2-dimensional affine group. Crucial to this reduction is the following

Lemma 29. *Let Γ be a subgroup of $\mathrm{Upp}_d(\mathbb{C})$. If Γ is not virtually nilpotent, then there is a homomorphism $\rho : \Gamma \to \mathrm{Aff}(\mathbb{C})$, whose image is not virtually nilpotent.*

Proof. We prove the result for subgroups Γ of a connected solvable algebraic group $\mathbb{G}$ in place of $\mathrm{Upp}_d(\mathbb{C})$. (This is equivalent to our assumption by the Lie-Kolchin theorem [30, 17.6]). We will work by induction on $\dim \mathbb{G}$. Without loss of generality (passing to the Zariski closure G of Γ), we may assume that $\Gamma = G$ is Zariski-closed, because if $\rho(\Gamma)$ is virtually nilpotent so will be $\rho(G)$. Under the assumption that G is Zariski-connected, a proof of this lemma can be found in [7, Lemma 10.7]. We need some adjustments to handle the general case.

To prove Lemma 29, we have to find a character $\chi : G \to \mathbb{C}^*$, and a non-trivial cocycle $\beta : G \to \mathbb{C}$, that is a map such that $\beta(gh) = \beta(g) + \chi(g)\beta(h)$, and $\beta(\ker \chi) \neq 0$. Then the map $G \to \mathrm{Aff}(\mathbb{C})$ sending g to the matrix

$$\begin{pmatrix} \chi(g) & \beta(g) \\ 0 & 1 \end{pmatrix}$$

gives the desired homomorphism.

Since $[\mathbb{G}, \mathbb{G}]$ is unipotent, we can apply Lemma 27 above to G. Hence there is a diagonalizable subgroup H of G such that $G = HG_u$. The subgroup H lies in a maximal torus of $\mathbb{G}$, say T, so that $\mathbb{G} = T \cdot U$, where U is the unipotent radical of $\mathbb{G}$. Note further that $G_u = G \cap U$.

Let Z be the center of U. It is a normal algebraic subgroup of $\mathbb{G}$ of positive dimension. If G° acts trivially on Z by conjugation, then we may

pass to $\mathbb{G}/Z$ and apply induction, since then the image of G in $\mathbb{G}/Z$ will not be virtually nilpotent.

So assume that its action is not trivial. Note that the G-action on Z factors through $\mathbb{G}/U \simeq T$. Since T is a torus, its action on the additive group $Z \simeq \mathbb{C}^d$ splits into weight spaces. There is a weight $\chi : T \to \mathbb{C}^*$ such that $\chi(H^\circ) \neq 1$. Let $Z_\chi \leqslant Z$ be a one-dimensional subspace in the eigenspace of χ, so that $tzt^{-1} = \chi(t)z$ when $t \in T$.

Note that Z_χ is a normal subgroup of $\mathbb{G}$. We can assume that H° acts trivially on U/Z_χ, for otherwise the image of G in $\mathbb{G}/Z_\chi$ would not be virtually nilpotent and we could again use induction. This means that $[H^\circ, U] \leqslant Z_\chi$.

On the other hand H° does not commute with G_u, for otherwise G° would be nilpotent. Pick $h_0 \in H^\circ$ such that $[h_0, G_u] \neq 1$. Since $G = HG_u$, and H commutes with h_0, we see that $[h_0, G] \leqslant Z_\chi$. Then we set $\beta(g) = [h_0, g]$ for $g \in G$, after identifying Z_χ with the additive group of $\mathbb{C}$. This yields the desired a non trivial cocycle as claimed and ends the proof of the lemma. $\square$

Remark 30. If $\mathbb{G}$ is a K-split connected solvable K-subgroup of $\mathrm{GL}_d(\mathbb{C})$, where K is some subfield of $\mathbb{C}$ and $G \leqslant \mathbb{G}$ is a closed algebraic K-subgroup, which is not virtually nilpotent, then replacing everywhere *maximal torus* by *K-split maximal torus*, the proof above combined with Remark 28 shows that the homomorphism $\rho : G \to \mathrm{Aff}(\mathbb{C})$ we have constructed is defined over K.

Recall M_λ denotes the Mahler measure of λ and that we have adopted the convention that $M_\lambda = 2$ if λ is transcendental. We conclude:

Corollary 31. *Let S be a finite subset of $\mathrm{Upp}_d(\mathbb{C})$ containing the identity. Assume that $\langle S \rangle$ is not virtually nilpotent, then there is $\lambda \in \mathbb{C}^\times$, which is not a root of unity, such that*

$$\rho_S \geqslant \frac{1}{9} \log M_\lambda.$$

Proof. This follows from the combination of Lemma 29 and Corollary 26. $\square$

Remark 32. If K_S denotes the field generated by the matrix entries of each element $s \in S$, then λ in the previous statement can be found in $K_S \setminus \{0\}$. This follows from the same argument together with Remarks 28 and 30.

To handle virtually solvable subgroups not necessarily contained in $\mathrm{Upp}_d(\mathbb{C})$, we need the following lemma.

Lemma 33. *Let G be a group, and H a subgroup with $G = SH$ for some finite generating subset S of G containing 1 (but not necessarily symmetric). Let $\rho : H \to \mathrm{Aff}(\mathbb{C})$ be a homomorphism with non virtually nilpotent image. Then the subgroup generated by $S^3 \cap H$ has a non virtually nilpotent image under ρ.*

Proof. Let χ be the character $H \to \mathbb{C}^*$ induced by the natural homomorphism $\mathrm{Aff}(\mathbb{C}) \to \mathbb{C}^*$. First we claim that $\chi(\langle S^3 \cap H \rangle)$ is infinite. Indeed H is generated by the elements of the form $s_1 s_2 s_3^{-1}$ belonging to H, with each s_i in S. At least one of them must map to an element of infinite order under χ, say $\chi(s_1 s_2 s_3^{-1})$ has infinite order. Let $s_4 \in S$ be such that $s_3^{-1} \in s_4 H$.

Then $s_3 s_4 \in H$. So we see that either $\chi(s_3 s_4)$ has infinite order, or else $s_1 s_2 s_4 = s_1 s_2 s_3^{-1} s_3 s_4$ has infinite order. This proves the claim.

So pick $\gamma \in S^3 \cap H$ with $\chi(\gamma)$ of infinite order. Now observe that every virtually nilpotent subgroup of $\mathrm{Aff}(\mathbb{C})$ containing $\rho(\gamma)$ must be abelian. Indeed if it contains an element not commuting with $\rho(\gamma)$, then the commutator will be a non-trivial translation t, but the subgroup generated by $\rho(\gamma)^n$ and t^n is nilpotent for no $n \in \mathbb{N}$ (the centralizers of these two elements have trivial intersection, so there is no center).

So if $\rho(\langle S^3 \cap H \rangle)$ were virtually nilpotent, it would be abelian. However $\rho(H)$ is not abelian, and it is generated by the $\rho(s_1 s_2 s_3^{-1})$ with $s_1 s_2 s_3^{-1}$ belonging to H and $s_1, s_2, s_3 \in S$. The centralizer of $\rho(\gamma)$ in $\mathrm{Aff}(\mathbb{C})$ is abelian. Pick such elements with $[\rho(s_1 s_2 s_3^{-1}), \rho(\gamma)] \neq 1$. As above let $s_4 \in S$ with $s_3 s_4 \in H$. We see that either $\rho(s_3 s_4)$ does not commute with $\rho(\gamma)$, or else $\rho(s_1 s_2 s_4)$ does not commute with $\rho(\gamma)$. In both cases $\rho(\langle S^3 \cap H \rangle)$ is not abelian, hence not virtually nilpotent. The lemma is proved. $\qquad\square$

We can now conclude the

Proof of Theorem 20. Let G be the Zariski closure of the subgroup Γ generated by S. Let T be a maximal torus of G°. For every $g \in G$, gTg^{-1} is a maximal torus, hence is conjugate to T by an element of G°. This shows that $G = N_G(T)G^\circ$.

Diagonalizing T, we see that the centralizer $Z_G(T)$ of T in G has index at most $d!$ in $N_G(T)$. By the result of Platonov mentioned at the beginning of the proof of Lemma 27, there is a finite subgroup F such that $Z_G(T) = FZ_G(T)^\circ$, and Jordan's theorem implies that there is an abelian subgroup A in F with index at most $J(d)$.

Now note that $AG^\circ = ATG_u$, where G_u is the unipotent radical of G, and AT is abelian. Consequently the commutator subgroup of the subgroup AG° is unipotent. So Lemma 27 shows that AG° can be conjugated inside $\mathrm{Upp}_d(\mathbb{C})$ by an element of $\mathrm{GL}_d(\mathbb{C})$. Let H be the Zariski-closure of $\Gamma \cap AG^\circ$.

Since $H \cap \Gamma$ has finite index in Γ, it is not virtually nilpotent. Now by Lemma 29 there is a homomorphism $\rho : H \to \mathrm{Aff}(\mathbb{C})$ with non virtually nilpotent image. In fact H has index at most $d!J(d)$ in Γ, so setting $S_0 = S^{d!J(d)}$, we see that $S_0 H = \Gamma$, and we may apply Lemma 33 to conclude that $\langle S_0^3 \cap H \rangle$ has non virtually nilpotent image under ρ. Hence Corollary 26 shows that $\rho_{S_0^3 \cap H} \geqslant \frac{1}{9} M_\lambda$ for some $\lambda \neq 0$ not a root of unity. The result follows immediately since $\rho_S \geqslant \frac{1}{3d!J(d)} \rho_{S_0^3}$. $\qquad\square$

Remark 34. We note that if K_S denotes the field generated by the matrix entries of each element $s \in S$, then λ in Theorem 20 can be found in a finite extension of degree at most $d!$ over K_S. To see this we only need to keep track of the field of definition at every step in the previous argument. In brief, by Remark 28 the group F can be chosen in $G(K_S)$, then the Zariski-closure H of $\Gamma \cap AG^\circ$ will be defined over K_S and can be triangularized by some element of $\mathrm{GL}_d(\mathbb{C})$. It follows that there is a field extension $\widehat{K}_S$ of K_S with degree at most $d!$ such that H can be triangularized by an element of $\mathrm{GL}_d(\widehat{K}_S)$. From Remark 30 the homomorphism $\rho : H \to \mathrm{Aff}(\mathbb{C})$ given by Lemma 29 will then be defined over $\widehat{K}_S$. The rest of the proof invoking

Lemma 33 and Corollary 26 is identical, except we make use of Remark 32 to guarantee that λ belongs to $\widehat{K}_S^\times$.

Proof of Corollary 24. From Theorem 20 and the previous remark, we have $\rho_S \geqslant \frac{1}{27d!J(d)} \log M_\lambda$. for some λ with degree at most $d!$ over $\mathbb{Q}$. The conclusion follows then easily from Dobrowolski's lower bound on the Mahler measure of algebraic numbers [18] (in fact a much weaker bound is enough) and the cited bounds for $J(d)$ (see Remark 22). $\qquad\square$

5. Further directions and open problems

We already mentioned in the introduction that given $\varepsilon > 0$ there exist a finitely generated group $\Gamma = \langle S \rangle$ such that $0 < \rho_S < \varepsilon$.

These examples of groups with slow exponential growth were constructed by Grigorchuk and de la Harpe in [24] out of a presentation for the Grigorchuk group of intermediate growth. They are virtually a product of finitely many free groups. Taking a suitable quotient such groups can be made solvable, in fact even metabelian-by-(finite 2-group), as was shown in [4]. However the solvability class in these examples is not bounded.

We record here the following:

Problem 1. *Given $r \in \mathbb{N}$, is there $c_r > 0$ such that if S is a finite generating subset of a solvable group with solvability class bounded by r, then either $\rho_S = 0$ and $\langle S \rangle$ is virtually nilpotent, or $\rho_S > c_r$?*

In view of Lemma 16 (or [12, §7]) a positive answer to this question implies the Lehmer conjecture. It would be nice to investigate, as we did in this paper for linear groups, whether the converse holds as well. To that end we state the following version of the previous problem.

Problem 2. *Given $r \in \mathbb{N}$, is there $c_r > 0$ such that if S is a finite generating subset of a solvable group with solvability class bounded by r, then either $\rho_S = 0$ and $\langle S \rangle$ is virtually nilpotent, or there is a number $\lambda \in \overline{\mathbb{Q}}^\times$ not a root of unity such that*

$$\rho_S > c_r \log M_\lambda?$$

The problem obviously reduces to the case when the group is just not virtually nilpotent in the sense that every proper quotient of the group is virtually nilpotent. Such groups are known to be virtually metabelian [12, 25]. Those that are metabelian (i.e. $r = 2$) embed in $\mathrm{Aff}(K)$ for some field K, hence for those the answer to the above problem is positive and the proof is easy. However to handle to case when $r > 2$, one needs new ideas to overcome the finite index issue.

Another interesting question is whether there are some numbers λ such that $h_\lambda = \log M_\lambda$. In particular, the case of Salem numbers would be very interesting, because in that case $h_\lambda = \log M_\lambda = \log \lambda^{-1}$ is equivalent to $\dim \mu_\lambda = 1$ by (1.2).

Problem 3. *Is it true that $h_\lambda = \log \lambda^{-1}$ for all Salem numbers $\lambda \in (1/2, 1)$?*

Observe that Salem numbers have conjugates on the unit circle, hence Proposition 7 does not apply. However, we learnt from Paul Mercat [36] that there are examples outside the scope of Proposition 7, such that $h_\lambda <$

$\log M_\lambda$. Such examples are the roots of the polynomial $x^6 + x^5 + x^4 - x^3 + x^2 + x + 1$, which are three pairs of complex conjugates that are inside, on and outside the unit circle, respectively, and $M_\lambda < 2$. Mercat showed that $\rho_\lambda < \log M_\lambda$ by computing the first 8 steps of the random walk and finding that $|\operatorname{Supp}(\mu_\lambda^{(8)})| < M_\lambda^8$.

REFERENCES

[1] J. C. Alexander and D. Zagier. The entropy of a certain infinitely convolved Bernoulli measure. *J. London Math. Soc. (2)* 44 (1):121–134, 1991. ↑5

[2] F. Amoroso and S. David. Le problème de Lehmer en dimension supérieure. *J. Reine Angew. Math.* 513:145–179, 1999. ↑5

[3] C. Bandt. On the Mandelbrot set for pairs of linear maps. *Nonlinearity* 15 (4):1127–1147, 2002. ↑4

[4] L. Bartholdi and Y. de Cornulier. Infinite groups with large balls of torsion elements and small entropy. *Arch. Math. (Basel)* 87 (2):104–112, 2006. ↑30

[5] A. Borel and J.-P. Serre. Théorèmes de finitude en cohomologie galoisienne. *Comment. Math. Helv.* 39:111–164, 1964. ↑26

[6] P. Borwein, E. Dobrowolski, and M. J. Mossinghoff. Lehmer's problem for polynomials with odd coefficients. *Ann. of Math. (2)* 166 (2):347–366, 2007. ↑5

[7] E. Breuillard and T. Gelander. A topological Tits alternative. *Ann. of Math. (2)* 166 (2):427–474, 2007. ↑27

[8] _____. Uniform independence in linear groups. *Invent. Math.* 173 (2):225–263, 2008. ↑3

[9] E. Breuillard and P. P. Varjú. *On the dimension of Bernoulli convolutions*, 2016. Work in progress. ↑2

[10] E. Breuillard. Diophantine geometry and uniform growth of finite and infinite groups. *Proceedings of the International Congress of Mathematicians, 2014 ICM (Seoul).* ↑2

[11] _____. A strong tits alternative. *arXiv:0804.1395.* ↑3, 24

[12] _____. On uniform exponential growth for solvable groups. *Pure Appl. Math. Q.* 3 (4, Special Issue: In honor of Grigory Margulis. Part 1):949–967, 2007. ↑3, 23, 30

[13] _____. A height gap theorem for finite subsets of $GL_d(\overline{\mathbb{Q}})$ and nonamenable subgroups. *Ann. of Math. (2)* 174 (2):1057–1110, 2011. ↑3

[14] _____. Heights on SL_2 and free subgroups. In Geometry, rigidity, and group actions, pages 455–493. 2011. ↑3, 24

[15] M. J. Collins. On Jordan's theorem for complex linear groups. *J. Group Theory* 10 (4):411–423, 2007. ↑23

[16] T. M. Cover and J. A. Thomas. *Elements of information theory*. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, Second, 2006. ↑9

[17] J. D. Dixon. *The structure of linear groups*. Van Nostrand-Reinhold, 1971. Van Nostrand Reinhold mathematical studies, Volume 37. ↑23

[18] E. Dobrowolski. On a question of Lehmer and the number of irreducible factors of a polynomial. *Acta Arith.* 34 (4):391–401, 1979. ↑24, 30

[19] P. Erdős. On a family of symmetric Bernoulli convolutions. *Amer. J. Math.* 61:974–976, 1939. ↑1

[20] A. Eskin, S. Mozes, and H. Oh. On uniform exponential growth for linear groups. *Invent. Math.* 160 (1):1–30, 2005. ↑3

[21] D.-J. Feng and H. Hu. Dimension theory of iterated function systems. *Comm. Pure Appl. Math.* 62 (11):1435–1500, 2009. ↑22

[22] A. M. Garsia. Arithmetic properties of Bernoulli convolutions. *Trans. Amer. Math. Soc.* 102:409–432, 1962. ↑21

[23] _____. Entropy and singularity of infinite convolutions. *Pacific J. Math.* 13:1159–1169, 1963. ↑4, 5, 18, 22

[24] R. Grigorchuk and P. de la Harpe. Limit behaviour of exponential growth rates for finitely generated groups. In Essays on geometry and related topics, Vol. 1, 2, pages 351–370. 2001. ↑2, 30

[25] J. R. J. Groves. Soluble groups with every proper quotient polycyclic. *Illinois J. Math.* 22 (1):90–95, 1978. ↑30

[26] K. G. Hare and N. Sidorov. A lower bound for Garsia's entropy for certain Bernoulli convolutions. *LMS J. Comput. Math.* 13:130–143, 2010. ↑5

[27] M. Hochman. On self-similar sets with overlaps and inverse theorems for entropy. *Ann. of Math. (2)* 180 (2):773–822, 2014. ↑1, 3, 6, 21

[28] A. Horn. On the eigenvalues of a matrix with prescribed singular values. *Proc. Amer. Math. Soc.* 5:4–7, 1954. ↑15

[29] R. A. Horn and C. R. Johnson. *Topics in matrix analysis.* Cambridge University Press, Cambridge, 1994. Corrected reprint of the 1991 original. ↑15

[30] J. E. Humphreys. *Linear algebraic groups.* Springer-Verlag, New York-Heidelberg, 1975. Graduate Texts in Mathematics, No. 21. ↑25, 26, 27

[31] K. I. and M. M. Sumset and inverse sumset inequalities for differential entropy and mutual information. *preprint arXiv:1206.0489*, 2012. ↑5, 11

[32] C. Jordan. Mémoire sur les équations différentielles linéaires à intégrale algébrique. *J. Reine Angew. Math.* 84:89–215, 1878. ↑23

[33] V. A. Kaĭmanovich and A. M. Vershik. Random walks on discrete groups: boundary and entropy. *Ann. Probab.* 11 (3):457–490, 1983. ↑11

[34] S. P. Lalley. β-expansions with deleted digits for Pisot numbers β. *Trans. Amer. Math. Soc.* 349 (11):4355–4365, 1997. ↑5

[35] M. Madiman. On the entropy of sums. *In Information Theory Workshop, 2008. ITW 08. IEEE*:303–307, 2008. ↑5, 11

[36] P. Mercat. Private communication. ↑30

[37] ———. Semi-groupes fortement automatiques. *Bull. Soc. Math. France* 141 (3):423–479, 2013. ↑5

[38] M. Mignotte. Sur les multiples des polynômes irréductibles. *Bull. Soc. Math. Belg.* 27 (3):225–229, 1975. ↑4

[39] M. Pathiaux. Sur les multiples de polynômes irréductibles associés à certains nombres algébriques. In Séminaire Delange-Pisot-Poitou, 14e année (1972/73), Théorie des nombres, Fasc. 1, Exp. No. 13, pages 9. 1973. ↑4

[40] Y. Peres, W. Schlag, and B. Solomyak. Sixty years of Bernoulli convolutions. In Fractal geometry and stochastics, II (Greifswald/Koserow, 1998), pages 39–65. 2000. ↑1

[41] J. Peters. Entropy on discrete abelian groups. *Adv. in Math.* 33 (1):1–13, 1979. ↑6

[42] V. P. Platonov. The theory of algebraic linear groups and periodic groups. *Izv. Akad. Nauk SSSR Ser. Mat.* 30:573–620, 1966. ↑26

[43] M. S. Raghunathan. *Discrete subgroups of Lie groups.* Springer-Verlag, New York-Heidelberg, 1972. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 68. ↑23

[44] R. Salem. A remarkable class of algebraic integers. Proof of a conjecture of Vijayaraghavan. *Duke Math. J.* 11:103–108, 1944. ↑2

[45] P. Shmerkin. On the exceptional set for absolute continuity of Bernoulli convolutions. *Geom. Funct. Anal.* 24 (3):946–958, 2014. ↑1

[46] T. Tao. Sumset and inverse sumset theory for Shannon entropy. *Combin. Probab. Comput.* 19 (4):603–639, 2010. ↑5, 11

[47] T. Tao and V. Vu. *Additive combinatorics.* Cambridge Studies in Advanced Mathematics, vol. 105. Cambridge University Press, Cambridge, 2006. ↑11

[48] P. P. Varjú. *Absolute continuity of Bernoulli convolutions for algebraic parameters*, 2016. arXiv:1602.00261v1. ↑4

[49] B. A. F. Wehrfritz. *Infinite linear groups. An account of the group-theoretic properties of infinite groups of matrices.* Springer-Verlag, New York-Heidelberg, 1973. Ergebnisse der Matematik und ihrer Grenzgebiete, Band 76. ↑23, 26, 27

Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, UK

E-mail address: emmanuel.breuillard@maths.cam.ac.uk

Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, UK

E-mail address: pv270@dpmms.cam.ac.uk