HEIGHTS ON SL_2 AND FREE SUBGROUPS

EMMANUEL BREUILLARD

ABSTRACT. In this mostly expository paper, we discuss the strong uniform Tits Alternative and give a complete proof of it in the special case of $GL_2(\mathbb{C})$. The main arithmetic ingredient, the height gap theorem, is also given a complete treatment in that case. We then prove several applications involving expansion properties of $SL_d(\mathbb{Z}/p\mathbb{Z})$, a uniform l^2 spectral gap, and diophantine properties of subgroups of $GL_d(\mathbb{C})$.

CONTENTS

1. INTRODUCTION

1.1. **Statement of the main results.** Recall that the Tits alternative [31] asserts that any finitely generated subgroup of $GL_d(K)$, where K is some field, contains a non abelian free subgroup on two generators unless it is amenable, or equivalently in this case, unless it contains a solvable subgroup of finite index (i.e. is virtually solvable). In [7] and [8], we showed the following uniform version of Tits' theorem :

Theorem 1.1. (**strong uniform Tits Alternative** [8]) For every $d \in \mathbb{N}$ there is $N = N(d) \in \mathbb{N}$ such that if K is any field and F a finite symmetric subset of $GL_d(K)$ containing 1, either F^N contains two elements which freely generate

Date: November 20th, 2007
The author acknowledges support from the IAS Princeton through NSF Grant DMS-0635607.

a non abelian free group, or the group generated by F is virtually solvable (i.e. contains a finite index solvable subgroup).

We have denoted by $F^n = F \cdot ... \cdot F$ the product set of n copies of F. In this paper we will discuss some consequences of this result and we will give a self-contained proof of it in the special case of SL_2 (hence equivalently GL_2) and K any field of characteristic 0 (note that this is equivalent to proving the result for $K = \mathbb{C}$, since every finitely generated field of characteristic 0 embeds in $\mathbb{C}$). This case is already representative of the general case as it captures the main difficulty, namely treat all number fields in a uniform way. The proof given in [8] is an elaboration of the proof for SL_2 that we are about to give.

Theorem 1.1 improves earlier refinements of Tits' theorem due to Eskin-Mozes-Oh (see [13]) and to T. Gelander and the author (see [9]). These two papers were concerned with the S-arithmetic version of Theorem 1.1, namely they proved uniformity of N (the "freeness radius" of F) for sets F with coefficients inside a fixed finitely generated ring. While in [13] the main concern was to prove uniform exponential growth by constructing generators of a free semigroup in F^N, in [9] it was shown that the result of [13] could be pushed to get generators of a free group in F^N, where N was depending only on d and on the ring generated by the matrix coefficients of the elements of F. Our main contribution in Theorem 1.1 is to remove the dependence on the ring of coefficients. As in Tits' proof or in [9], the proof of Theorem 1.1 can be divided into an arithmetic step on the one hand and a geometric step on the other. While in [9] (as well as in Tits' original theorem) the arithmetic step was the easier one and most of the work lied in showing that a certain geometric configuration (the so-called "ping-pong") did arise, roles are reversed in our proof of Theorem [8] and the arithmetic step is the harder step, while the geometric step routinely follows Tits' original proof after a careful check that all estimates are indeed uniform over all local fields. For SL_2 however none of the usual difficulties of higher rank arise and as will become clear below this geometric step is even more transparent (in that case or other rank 1 situations this geometric step can also be performed differently by acting directly on the hyperbolic space/tree as has been pointed out by T. Gelander).

The proof of Theorem 1.1 is effective in the sense that the constant N can, in principle, be made explicit. Examples due to Grigorchuk and de la Harpe [18] imply that $N(d)$ must tend to infinity with d. They exhibited a sequence $(\Gamma_n)_{n\geq 0}$ ($\Gamma_n \leq GL(n, \mathbb{Z})$) of 4-generated subgroups whose growth exponent decays to 1. These groups arise by chopping the usual presentation of the Grigorchuk group after finitely many relators (see [18]).

The arithmetic step in Theorem 1.1 relies on the following result, proved by the author in [7], which can be seen as a global adelic analogue of the Margulis Lemma about discrete subgroups of isometries in hyperbolic geometry. We will present here a self-contained proof of it in the case of SL_2.

Let $\overline{\mathbb{Q}}$ be an algebraic closure of $\mathbb{Q}$. In [7] we introduced the *arithmetic spectral radius* (or normalized height) of F, defined as

$$\widehat{h}(F) = \lim_{n \to +\infty} \frac{1}{n} h(F^n),$$

where h is the (absolute) *height* defined for F a finite subset of $M_d(\overline{\mathbb{Q}})$ by :

$$h(F) = \frac{1}{[K : \mathbb{Q}]} \sum_{v \in V_K} n_v \log^+ ||F||_v$$

where $\log^+ = \max\{0, \log\}$, K is the number field generated by the matrix coefficients of F, V_K is the set of all places of K, and $||F||_v = \max\{||f||_v, f \in F\}$ is the maximal operator norm of $f \in F$, where $||f||_v = \max_{x \neq 0} ||f(x)||_v/||x||_v$ for the standard norm $||x||_v$ induced on K_v^d by the standard absolute value $|\cdot|_v$ on the completion K_v of K associated to $v \in V_K$. We have also set $n_v = [K_v, \mathbb{Q}_v]$, where $\mathbb{Q}_v$ is the field of p-adic numbers if $v|p$ is finite and is $\mathbb{R}$ if v is infinite. The normalization of the absolute value $|\cdot|_v$ is such that $|\lambda|_v^{n_v}$ is the modulus of the multiplication by λ on K_v.

The quantities $h(F)$ and $\widehat{h}(F)$ are well defined, i.e. they are independent of the chosen number field. Moreover $\widehat{h}(F)$ is invariant under conjugation by elements from $GL_d(\overline{\mathbb{Q}})$. The main statement is:

Theorem 1.2. (**Height Gap Theorem** [7]) There is a positive constant $\varepsilon = \varepsilon(d) > 0$ such that if F is a finite subset of $GL_d(\overline{\mathbb{Q}})$ generating a non virtually solvable subgroup Γ, then

$$\widehat{h}(F) > \varepsilon.$$

Moreover, if the Zariski closure of Γ is semisimple, then

$$\widehat{h}(F) \leq \inf_{g \in GL_d(\overline{\mathbb{Q}})} h(gFg^{-1}) \leq C \cdot \widehat{h}(F)$$

for some absolute constant $C = C(d) > 0$.

When $d = 2$, i.e. for SL_2, the first part of Theorem 1.2 has the following geometric interpretation: either there is a finite place v where F acts without global fixed point on the corresponding Bruhat-Tits tree, or after applying some Galois automorphism of $\mathbb{C}$, the set F, as it acts on the hyperbolic 3-space $\mathbb{H}^3$ via $SL_2(\mathbb{C})$, moves every point away from itself by a positive absolute constant ε. This is analogous to the Margulis lemma in hyperbolic geometry, according to which if F generates a discrete subgroup of $SL_2(\mathbb{C})$ which is not virtually nilpotent, then every point of $\mathbb{H}^3$ is moved away from itself by some element of F by some fixed constant (see [30]).

The main purpose of Theorem 1.2 is to yield in F, or a bounded power of F, a nice hyperbolic element, i.e. a semisimple matrix whose eigenvalue is of modulus at least 2, say, in some local completion.

Finally we point out that our proof of Theorem 1.1 yields a free subgroup with the extra property that it is *uniformly undistorted* in the original subgroup, namely:

Theorem 1.3. (**Uniformly quasi-isometrically embedded free subgroup** [8]). There is a constant $C = C(d) > 0$ such that if K is any field and Γ is any non virutally solvable subgroup of $GL_d(K)$ generated by a finite symmetric subset F (giving rise to a word metric $d_\Gamma(\cdot,\cdot)$ on Γ) there exists a free subgroup H of Γ generated by two elements (giving rise to a word metric $d_H(\cdot,\cdot)$) such that for all $h \in H$

$$\frac{1}{C \cdot |F|^C} \cdot d_\Gamma(1,h) \leq d_H(1,h) \leq d_\Gamma(1,h).$$

1.2. **Some consequences for uniform growth, spectral gap and diophantine properties.** We will prove here these corollaries for $GL_d(\mathbb{C})$, $d \geq 2$.

Corollary 1.4. (**Strong uniform exponential growth**) There is $\varepsilon = \varepsilon(d) > 0$ such that if F is a finite subset of $GL_d(\mathbb{C})$ containing 1 and generating a non amenable subgroup, then for all $n \geq 1$, $|F^n| \geq (1+\varepsilon)^n$. In particular,

$$\rho_F = \lim_{n\to+\infty} \frac{1}{n} \log|F^n| \geq \log(1+\varepsilon) > 0.$$

Let us remark that it may be the case that ρ_F is bounded away from 0 assuming only that F generates a non virtually nilpotent subgroup of $GL_2(\mathbb{C})$. However we observed in [6] that such an assertion, if true, *would imply the Lehmer conjecture* about the Mahler measure of algebraic numbers. We also observed there that although every linear solvable group of exponential growth contains a free semigroup, no analog of Theorem 1.1 (the UF property of [1]) holds for solvable groups, namely one may find sets F_n in $GL_2(\mathbb{C})$ containing 1 and generating a solvable subgroup of exponential growth, such that no pair of elements in $(F_n)^n$ may generate a free semigroup.

Von Neumann showed that groups containing a free subgroup are non amenable, i.e. have a spectral gap in ℓ^2. The uniformity in Theorem 1.1 implies also a uniformity for the spectral gap (see [29] for this observation). More precisely:

Corollary 1.5. (**Strong uniform Spectral Gap in** ℓ^2) There is $\varepsilon = \varepsilon(d) > 0$ with the following property. If F is a finite subset of $GL_d(\mathbb{C})$ containing the identity and generating a non amenable subgroup and if Γ is a countable subgroup of $GL_d(\mathbb{C})$ containing F and $f \in \ell^2(\Gamma)$, then there is $\sigma \in F$ such that

$$\sum_{x\in\Gamma} \left|f(\sigma^{-1}x) - f(x)\right|^2 \geq \varepsilon \cdot \sum_{x\in\Gamma} |f(x)|^2$$

In particular, if F in $GL_d(\mathbb{C})$ is a finite subset containing the identity and generating a non amenable subgroup, then for every finite subset A in $GL_d(\mathbb{C})$, we have $|FA| \geq (1+\varepsilon)|A|$.

In [24] Lubotzky, Phillips and Sarnak showed that for the compact Lie group $G = SU(2)$, the spectral measure of the "Hecke operators" $T_\mu = \frac{1}{2k}\sum_{1\leq i\leq k} g_i + g_i^{-1}$ acting on $\mathbb{L}_0^2(G)$ is supported on $[-m_\mu, m_\mu]$ where m_μ is the norm of T_μ viewed as an operator on $\ell^2(\Gamma)$, Γ being the abstract group generated by the g_i's. This spectral measure is by definition the limiting distribution of the eigenvalues of T_μ on the n^{th} dimensional representation of G. Corollary 1.5 implies that m_μ is bounded away from 1 independently of μ as soon as k is fixed and Γ is a non amenable subgroup of G. In other words, there is $\varepsilon = \varepsilon(k) > 0$ such that the proportion of eigenvalues of T_μ lying in $[-1, -1+\varepsilon] \cup [1-\varepsilon, 1]$ tends to 0 as n tends to infinity. The analogous result for Cayley graphs of $SL_2(\mathbb{Z}/p\mathbb{Z})$ is also a direct consequence of Corollary 1.5, i.e. the spectral measure of any limit of such Cayley graphs is supported on $[-1+\varepsilon, 1-\varepsilon]$. It is believed (*spectral gap conjecture,* see [24] [28]) that 1 is never an accumulation point of eigenvalues of T_μ for any given μ (or at least almost any in $SU(2)$) whose support generates a non amenable subgroup.

Since by Kesten's theorem m_μ is also the exponential rate of decay of the return probability (see [21]), we also have:

Corollary 1.6. (**Strong uniform decay of return probability***)* There is $\varepsilon = \varepsilon(k, d) > 0$ with the property that for any non amenable k-generated subgroup Γ of $GL_d(\mathbb{C})$ we have

$$\mathbb{P}(S_n = 1) \leq (1-\varepsilon)^n$$

for all $n \geq 1$, where S_n is the simple random walk on Γ.

The next corollary gives an upper bound on the co-growth of subgroups of $GL_d(\mathbb{C})$.

Corollary 1.7. (**Co-growth gap**) Given $m \in \mathbb{N}$, there is $n(m) > 0$ such that for every $n \geq n(m)$ the following holds: $F = \{a_1, ..., a_m\} \subset GL_d(\mathbb{C})$ generates a non virtually solvable subgroup, if and only if the number of elements w in the free group F_m of word length n such that $w(a_1, ..., a_m) = 1$ is at most $(2m - 1 - \frac{\varepsilon}{m^D})^n$. Here $\varepsilon, D > 0$ are constants depending on d only.

This result can be paraphrased by saying that non amenable subgroups of $GL_d(\mathbb{C})$ are very strongly non amenable, i.e. have few relations. This puts a purely group theoretical restriction on an abstract finitely generated group given in terms of generators and relations to admit an embedding in $GL_d(\mathbb{C})$.

The uniformity in Theorem 1.1 allows to reduce mod p and we obtain a statement giving a lower bound on the girth of subgroups of GL_d in positive characteristic:

Corollary 1.8. (**Large girth**) Given $k, d \geq 2$, there is $N, M \in \mathbb{N}$ and $\varepsilon_0, C > 0$ such that for every prime p and every field K of characteristic p and any finite subset F with k elements generating a subgroup of $GL_d(K)$ which contains no solvable subgroup of index at most M, then F^N contains two elements a, b such that $w(a,b) \neq 1$ in $GL_d(K)$ for any non trivial word w in F_2 of length at most $f(p) = C \cdot (\log p)^{\varepsilon_0}$.

It was conjectured in [16] that the statement of Corollary 1.8 holds for generating subsets F of $SL_2(\mathbb{F}_p)$ with $\varepsilon_0 = 1$.

Theorem 1.3 on the uniform QI-embedding of the free subgroup yields a uniform bound for the distortion of the subgroup of large girth. This in turn gives uniform expansion for subsets of say $GL_2(\mathbb{F}_p)$ lying in a ball of radius $\leq (\log p)^{\varepsilon_0}$. Namely:

Corollary 1.9. (**Uniform expansion for small sets**) Given $k, d \geq 2$, there is $N, M \in \mathbb{N}$ and $\varepsilon_0, C, \alpha, \beta > 0$ such that for every prime p and every field K of characteristic p and any finite symmetric subset F with k elements, containing 1 and generating a subgroup of $GL_d(K)$ which contains no solvable subgroup of index at most M, then for any subset $A \subset F^{C\cdot(\log p)^{\varepsilon_0}}$ there is $s \in F$ such that $|sA \triangle A| \geq \alpha |A|$. In particular $\mu_F^{*n}(e) \leq (1-\beta)^n$ for all $n \leq C \cdot (\log p)^{\varepsilon_0}$, where μ_F is the uniform probability measure on F.

If we could get $\varepsilon_0 = 1$ in the above corollary, then applying the argument of Bourgain and Gamburd ([4]) would give a proof that the family of all Cayley graphs of $SL_2(\mathbb{Z}/p\mathbb{Z})$ for varying p but with a fixed number of generators forms an expander family. See Lubotzky's book [23] for background material on expanders. It was also proved in [16] that a random d-regular Cayley graph of $GL_2(\mathbb{F}_p)$ has girth at least $(1 - o(1)) \log_{d-1}(p)$. Here we obtain $\varepsilon_0 = 2^{-10}$ for GL_2, which is quite far.

In the same vein, one obtains the following two weak form of "non-Liouvilleness" for subgroups of $GL_d(\mathbb{C})$. Let d be some Riemannian distance on $GL_2(\mathbb{C})$.

Corollary 1.10. *(***Short words are not simultaneously very Liouville***)* Given d, there is $N_0 \in \mathbb{N}$ and $\varepsilon_1 > 0$ with the following property. For every finite set $F \subset GL_d(\mathbb{C})$ generating a non virtually solvable subgroup, there is $\delta_0(F) > 0$ such that for every $\delta \in (0, \delta_0)$ there are two short words $a, b \in F^N$ such that $d(w(a,b), 1) \geq \delta$ for every reduced word w in the free group F_2 with length $\ell(w)$ at most $(\log \delta^{-1})^{\varepsilon_1}$.

In [19] Kaloshin and Rodnianski proved that for $G = SU(2) \leq SL_2(\mathbb{C})$ almost every pair $(a,b) \in G \times G$ satisfies $d(w(a,b), 1) \geq \exp(-C(a,b) \cdot \ell(w)^2)$ for all $w \in F_2 \backslash \{e\}$ and some constant $C(a,b) > 0$. Besides it is easy to see that if $a, b \in GL_2(\overline{\mathbb{Q}})$ then the pair (a,b) satisfies the stronger diophantine condition $d(w(a,b), 1) \geq \exp(-C(a,b) \cdot \ell(w))$. It is conjectured in [28] and [16], that this stronger condition also holds for almost every pair $(a,b) \in SU(2)$.

Our result also allows us to estimate the number of words of length $\leq n$ that fall in a shrinking neighborhood of 1 in $GL_d(\mathbb{C})$. More precisely,

Corollary 1.11. *(***Weak diophantine property***)* There are $\tau, \varepsilon_1, C > 0$ with the following property. For every $\{a, b\} \leq GL_d(\mathbb{C})$ which generates a non virtually solvable subgroup, there is $\delta_0(a, b) > 0$ such that for every $\delta \in (0, \delta_0)$ and every $n \leq C(\log \delta^{-1})^{\varepsilon_1}$, the proportion of elements w in the free group F_2 of word length n such that $d(w(a, b), 1) \leq \delta$ is at most $\exp(-\tau n)$.

In [15], Gamburd, Jacobson and Sarnak, showed for $G = SU(2)$ that if a pair $(a, b) \in G$ satisfies the conclusion of Corollary 1.11 with $\varepsilon_1 = 1$ and $C > C_0$ (for some explicit $C_0 > 0$) then (a, b) has a spectral gap on $\mathbb{L}^2(G)$. In [5], Bourgain and Gamburd showed that if a pair $(a, b) \in G$ satisfies the above condition with $\varepsilon_1 = 1$ and some $C = C(a, b) > 0$, perhaps small, then (a, b) has a spectral gap on $\mathbb{L}^2(G)$. This latter condition is automatically satisfied if (a, b) satisfies the stronger diophantine condition above, for instance if $(a, b) \in GL_2(\overline{\mathbb{Q}})$. Hence these pairs have a spectral gap. It is unknown whether there are (topologically generating) pairs with no spectral gap.

Remark 1.12. We emphasize here that all the constants in the above theorems and corollaries can be effectively computed. Only at one point in the proof do we use a compactness argument. This is in our proof of Lemma 2.1 (b). However this statement can be given an effective proof valid in $M_d(\mathbb{C})$ (available upon request).

Comment on the proof of Corollaries 1.8 to 1.11: Observe (see Corollary 9.2) that the condition on a finite subset $F = \{A, B\}$ of $GL_d(\mathbb{C})$ that it should generate an amenable (or equivalently virtually solvable) subgroup of $GL_d(\mathbb{C})$ is an algebraic one, as is the condition that all short words in A and B satisfy a relation of length at most n. Thus the statement of Theorem 1.1 can be read as a union of countably many assertions of first order logic. According to the "compactness theorem" from model theory, since each assertion holds for $\mathbb{C}$ it must also hold for an arbitrary field K of sufficiently large characteristic (depending on n). This readily gives this existence of some function $f(p)$ going to $+\infty$ with p in Corollary 1.8. To derive the bound $(\log p)^{\varepsilon_0}$, as well as the bounds $(\log \delta^{-1})^{\varepsilon_0}$ in Corollaries 1.10 and 1.11, we use a standard version of the effective Nullstellensatz due to Masser and Wustholz (see [26]).

1.3. Outline of the paper and of the proof of Theorems 1.2 and 1.1 for $SL_2(\mathbb{C})$.

In Section 3 we introduce our main objects, the height and normalized height of a finite set F of matrices and prove basic properties about them. One of the key properties, the comparison between $\widehat{h}$ and e, relies crucially on Section 2, which is devoted to the proof of a key lemma, *the spectral radius lemma.* This lemma says in substance that unless F can be conjugated in a bounded part of SL_2, one will find a short word with letters in F with a large eigenvalue. In Section 4 we prove the first part of Theorem 1.2 (the height gap). The proof makes crucial use of equidistribution properties of algebraic numbers of small height and in particular a result of Zhang (Theorem 4.9) and Bilu's equidistribution theorem for Galois orbits (Theorem 4.4 below). Using the Eskin-Mozes-Oh escape lemma (see Lemma 4.8 below) we first reduce to the 2-generated case, when $F = \{Id, A, B\}$ say. We can also assume that A is diagonal. By making local estimates at each place, and with the help of Bilu's theorem, we then show that $\widehat{h}(F)$ small implies that the heights of b_{11}, b_{22} and $b_{12}b_{21}$ are small. But as $b_{11}b_{22} - b_{12}b_{21} = 1$, Zhang's theorem quickly yields to a contradiction if $b_{12}b_{21} \neq 0$. So $b_{12}b_{21} = 0$ and F is made of upper or lower triangular matrices.

Theorem 1.1 is of purely algebraic nature and we begin its proof by showing that its validity over $\overline{\mathbb{Q}}$ implies its validity over $\mathbb{C}$. One then needs to exhibit a place v where one can *play ping-pong* on the projective line $\mathbb{P}^1(K_v)$ for the local field K_v, as in Tits' proof of his alternative. These "ping-pong players" will be the generators of the desired free subgroup. To achieve this, one needs to be able to conjugate F in $SL_2(K_v)$ in such a way that three conditions are satisfied. First the norm $||F||_v$ ought to be controlled (up to a fixed power) by the maximal eigenvalue say $|\lambda|_v$ of an element, say A, lying in F (or F^N for a bounded N). Second $|\lambda|_v$ should be large enough, i.e. bounded away from 1. And third, at least one element, say B, from F, or F^N, must send the eigenvectors of A far apart from one another with a distance controlled by some negative power of $||F||_v$. This criterion for ping-pong is explained in Section 6.

In Section 7 we show that a place v with these properties does exist. This is done in two steps, first (Section 5) we show that the minimal height $\widehat{h}(F)$ can be almost achieved (up to multiplicative and additive constants) by the ordinary height $h(F)$ after possibly conjugating F inside $SL_2(\overline{\mathbb{Q}})$. This is the second half of Theorem 1.2 : this step uses the estimates needed in the first part of Theorem 1.2 (i.e. the proof of the height gap). In a second step (Section 7), we use the product formula on $\mathbb{P}^1(\overline{\mathbb{Q}})$ to show that the distances between eigenvectors of A and their images under B are controlled in terms of $h(F)$, and hence $\widehat{h}(F)$, only. This implies the existence of a place v satisfying the first and third conditions. Theorem 1.2 ensures that v can be chosen to satisfy the second condition also.

Sections 8 and 9 are devoted to the applications.

2. Spectral Radius Lemma for several matrices

In this section we state and prove the crucial Lemma 2.1. It says that given a finite set of matrices with coefficients in a local field, one may always find a short word with letters in that finite set whose maximal eigenvalue is as large as the minimal norm of the finite set. Together with Proposition 2.5 it can be interpreted as an analog for several matrices of the classical spectral radius lemma relating the maximal eigenvalue and the rate of growth of the powers of a matrix. This lemma expresses in a condensed form an idea from a key proposition of the work of Eskin-Mozes-Oh where the concept of an almost algebra was used to essentially achieve the same goal. We emphasize here that getting an equality in part (a) of Lemma 2.1 as opposed to a mere inequality like in part (b) of the same lemma is absolutely crucial in the whole proof and in particular in Theorem 1.2.

Let k be a local field of characteristic 0. Let $\|\cdot\|_k$ be the standard norm on k^2, that is the canonical Euclidean (resp. Hermitian) norm if $k = \mathbb{R}$ (resp. $\mathbb{C}$) and the sup norm ($\|(x, y)\|_k = \max\{|x|_k, |y|_k\}$) if k is non Archimedean. We will also denote by $\|\cdot\|_k$ the operator norm induced on $M_2(k)$ by the standard norm $\|\cdot\|_k$ on k^2. Let Q be a bounded subset of matrices in $M_2(k)$. We set

$$\|Q\|_k = \sup_{g\in Q} \|g\|_k$$

and call it the *norm of* Q. Let $\overline{k}$ be an algebraic closure of k. It is well known (see Lang's Algebra [22]) that the absolute value on k extends to a unique absolute value on $\overline{k}$, hence the norm $\|\cdot\|_k$ also extends in a natural way to $\overline{k}^2$ and to $M_2(\overline{k})$. This allows to define the *minimal norm* of a bounded subset Q of $M_2(k)$ as

$$E_k(Q) = \inf_{x\in GL_2(\overline{k})} \left\|xQx^{-1}\right\|_k$$

We will also need to consider the *maximal eigenvalue of* Q, namely

$$\Lambda_k(Q) = \max\{|\lambda|_k,\ \lambda \in spec(q), q \in Q\}$$

where $spec(q)$ denotes the set of eigenvalues (the spectrum) of q in $\overline{k}$. Finally let $R_k(Q)$ be the *spectral radius* of Q

$$R_k(Q) = \lim_{n\to+\infty} ||Q^n||_k^{\frac{1}{n}}$$

These quantities are related to one another. The key property concerning them is given in the following assertion (which also holds in $M_d(k)$, $k \geq 2$, see [7]).

Lemma 2.1. (**Spectral Radius Lemma**) Let Q be a bounded subset of $M_2(k)$,

(a) if k is non Archimedean, then $\Lambda_k(Q^2) = E_k(Q)^2$.

(b) if k is Archimedean, there is a constant $c \in (0, 1)$ independent of Q, such that $\Lambda_k(Q^2) \geq c^2 \cdot E_k(Q)^2$.

Proof. We make use of the following easy lemmas.

Lemma 2.2. Let L be a field and Q a subset of $M_2(L)$ such that Q and Q^2 consist of nilpotent matrices. Then there is a basis (u, v) of L^2 such that $Qu = 0$ and $Qv \subset Lu$.

Proof. For any $A, B \in Q$, we have $A^2 = B^2 = (AB)^2 = 0$. It follows, unless A or B are zero, that $\ker A = \operatorname{Im} A$ and $\ker B = \operatorname{Im} B$. Also if $AB \neq 0$, we get $\ker B = \ker(AB) = \operatorname{Im}(AB) = \operatorname{Im} A$, while if $AB = 0$, then $\operatorname{Im} B = \ker A$. So at any case $\ker A = \operatorname{Im} A = \ker B = \operatorname{Im} B$. So we have proved that the kernels and images of non zero elements of Q coincide and are equal to some line Lu, say. Pick $v \in L^2 \backslash \{Lu\}$, then (u, v) forms the desired basis. □

Lemma 2.3. Let k be a local field with ring of integers $\mathcal{O}_k$ and uniformizer π. Let $A = (a_{ij}) \in M_2(\mathcal{O}_k)$ such that $trace(A)$ and $\det(A)$ belong to (π^2) and $a_{11}, a_{22}, a_{21} \in (\pi)$, while $a_{12} \in \mathcal{O}_k^\times$. Then $a_{21} \in (\pi^2)$.

Proof. We have $a_{12}a_{21} = a_{11}a_{22} - \det(A) \in (\pi^2)$, and $a_{12} \in \mathcal{O}_k^\times$, hence $a_{21} \in (\pi)^2$. □

When k is a non-archimedean local field, if a set $Id+Q$ in $SL_2(k)$ and its square have only eigenvalues very close to 1, then it must fix pointwise the 1-neighborhood of some point in the Bruhat-Tits tree of $SL_2(k)$. This is essentially the content of the next lemma.

Lemma 2.4. (small eigenvalues implies large fixed point set) Let k be a local field with ring of integers $\mathcal{O}_k$ and uniformizer π together with an absolute value $|\cdot|_k$, which is (uniquely) extended to an algebraic closure $\overline{k}$ of k. Let Q be a subset of $M_2(\mathcal{O}_k)$ such that $\Lambda_k(Q)$ and $\Lambda_k(Q^2)$ are both $\leq |\pi|_k^3$. Then there is $T \in GL_2(k)$ such that $TQT^{-1} \subset \pi M_2(\mathcal{O}_k)$.

Proof. We can write Q as the disjoint union $Q_1 \cup \pi Q_2$ where Q_1 does not intersect $\pi M_2(\mathcal{O}_k)$. Let $Q' = Q_1 \cup Q_2$. Then $\Lambda_k(Q')$ and $\Lambda_k(Q'^2)$ are both $\leq |\pi|_k$. Hence projecting to $M_2(L)$, where L is the residue field $L = \mathcal{O}_k/(\pi)$, the matrices from Q' and Q'^2 become nilpotent. According to Lemma 2.2, one may find a basis $(\overline{u}, \overline{v})$ of L^2 such that $Q'\overline{u} = 0$ and $Q'\overline{v} \subset L\overline{u}$. According to Nakayama's lemma, this basis is the projection of a basis (u, v) of $\mathcal{O}_k^2$. Up to conjugating by a matrix in $GL_2(\mathcal{O}_k)$, we may assume that (u, v) is the canonical basis of $\mathcal{O}_k^2$. Therefore Q' consists of matrices $A = (a_{ij}) \in M_2(\mathcal{O}_k)$ with $a_{11}, a_{22}, a_{21} \in (\pi)$. Moreover, matrices in Q_1 satisfy $a_{12} \in \mathcal{O}_k^\times$ and hence by Lemma 2.3, $a_{21} \in (\pi^2)$. But for the matrices coming from πQ_2 also we have $a_{21} \in (\pi^2)$. So we have $a_{21} \in (\pi^2)$ for all matrices in Q. Let $T = diag(\pi, 1) \in GL_2(k)$. Then clearly $TQT^{-1} \subset \pi M_2(\mathcal{O}_k)$. □

We go back to the proof of Lemma 2.1. We first prove (b). By contradiction, if such a c did not exist, then we may find a sequence of such Q_n such that $\frac{\Lambda_k(Q_n^2)}{E_k(Q_n)^2} \to 0$. We can change Q_n into $\frac{Q_n}{E_k(Q_n)}$ and thus obtain a sequence of compact sets in $M_2(k)$ such that $E_k(Q_n) = 1$ with $\Lambda_k(Q_n^2) \to 0$ and $\Lambda_k(Q_n) \to 0$. and passing to a limit, we obtain a compact subset Q of $M_2(k)$ such that $\Lambda_k(Q^2) = \Lambda_k(Q) = 0$ while $E_k(Q) = 1$. By Lemma 2.2, we can thus find a basis (u, v) where $Qu = 0$ and $Qv \subset Lu$. But then conjugating Q by a suitable diagonal matrix can shrink the norm of Q as much as we want, hence $E_k(Q) = 0$. A contradiction.

We now prove (a). Let π be a uniformizer for k. Let $x = \log E_k(Q)$ where the log is taken in base $|\pi|_k^{-1}$. Suppose that $\Lambda_k(Q^2) < E_k(Q)^2$ and let $\varepsilon = x - \frac{1}{2}\log \Lambda_k(Q^2) > 0$. Then as $\Lambda_k(Q) \leq \Lambda_k(Q^2)^{\frac{1}{2}}$, we have $x - \log \Lambda_k(Q) \geq \varepsilon > 0$.

Note that with our choice of normalization, $\log \Lambda_k(Q^2) \in \frac{1}{2}\mathbb{Z}$. Let $n \in \mathbb{N}$ such that $2n\varepsilon > 3$. Let k_0 be the extension $k(\sqrt[2n]{\pi})$ where $\sqrt[2n]{\pi}$ is some $2n$-root of π in $\overline{k}$. Since $x = \log E_k(Q) = \log \inf\{||gQg^{-1}||_k, g \in GL_2(\overline{k})\}$, we may assume after possibly conjugating Q inside $GL_2(k_1)$, for some finite extension k_1 of k_0, that $y := \log ||Q||_k \leq x + \frac{1}{2n}$ and also that $y = \min\{\log ||gQg^{-1}||_k,\ g \in GL_2(k_1)\}$. Let π_1 be a uniformizer in k_1. Then $\log |\pi_1|_k^{-1} \leq \frac{1}{2n}$ and $y = \log |\pi_1|^m$ for some $m \in \mathbb{Z}$. Let $Q_y = \pi_1^{-m} Q \subset M_2(\mathcal{O}_{k_1})$. We get $\log E_k(Q_y) = x - y \leq 0$ and $\log \Lambda_k(Q_y) \leq \frac{1}{2}\log \Lambda_k(Q_y^2) = x - y - \varepsilon \leq -\varepsilon \leq -\frac{3}{2n} \leq \log |\pi_1|_k^3$. We are thus in a position to apply Lemma 2.4, which implies that Q_y, and hence Q itself, can be further conjugated inside $GL_2(k_1)$ so as to strictly reduce its norm. But this contradicts the minimality of y. □

Proposition 2.5. Let Q be a bounded subset of $M_2(k)$. We have

$$R_k(Q) = \lim_{n\to+\infty} E_k(Q^n)^{\frac{1}{n}} = \inf_{n\in\mathbb{N}} E_k(Q^n)^{\frac{1}{n}} = \lim_{n\to+\infty} \Lambda_k(Q^{2n})^{\frac{1}{2n}} = \sup_{n\in\mathbb{N}} \Lambda_k(Q^n)^{\frac{1}{n}}$$

Moreover if k is non Archimedean, $R_k(Q) = E_k(Q)$, while if k is Archimedean, then $c \cdot E_k(Q) \leq R_k(Q) \leq E_k(Q)$, where c is the constant from Lemma 2.1 (b).

Proof. We omit the proof: these identities follow either directly from the definitions or as a straightforward application of Lemma 2.1. □

Note that some periodicity phenomenon may arise if $Id \notin Q$, namely it may be that $\Lambda_k(Q^{2n+1}) = 1$ for all n while $\Lambda_k(Q^{2n})$ tends to infinity (for instance take for Q a set of symmetries around several points on a given geodesic in the hyperbolic plane). However if $Id \in Q$, then we do have $\lim_{n\to+\infty} \Lambda_k(Q^n)^{\frac{1}{n}} = R_k(Q)$.

Note also that if Q belongs to $SL_2(k)$, then $E_k(Q) \geq R_k(Q) \geq \Lambda_k(Q) \geq 1$ and all three quantities remain unchanged if we add Id to Q. The following lemma explains what happens if these quantities are close or equal to 1.

Lemma 2.6. *(*Linear growth of displacement squared*)* Suppose k is Archimedean (i.e. $k = \mathbb{R}$ or $\mathbb{C}$). Then we have for every $n \in \mathbb{N}$ and every bounded symmetric subset Q of $SL_2(k)$ containing Id,

$$E_k(Q^n) \geq E_k(Q)^{\sqrt{\frac{n-1}{8}}}. \tag{1}$$

Moreover,

$$\log R_k(Q) \geq c_1 \cdot \log E_k(Q) \cdot \min\{1, \log E_k(Q)\}$$

for some constant $c_1 > 0$. In particular $E_k(Q) = 1$ iff $R_k(Q) = 1$.

Proof. We use non-positive curvature of hyperbolic space $\mathbb{H}^3$. For $x \in \mathbb{H}^3$ set $L(Q, x) = \max_{g\in Q^n} d(gx, x)$ and $L(Q) = \inf_x L(Q, x)$. Fix $\varepsilon > 0$ and let $x_k \in \mathbb{H}^3$ be a point almost minimizing the displacement of Q^k, i.e. $L(Q^k, x_k) - \varepsilon \leq r_k = L(Q^k)$. Note that $r_k = 2\log E(Q^k)$. For each $g \in Q$, both $Q^{k-2}gx_k$ and $Q^{k-2}x_k$ lie in the intersection of the balls of radius $r_k + \varepsilon$ centered at x_k and at gx_k. By the CAT(0) inequality for the median, this intersection is contained in the ball of squared radius $(r_k+\varepsilon)^2-\frac{1}{4}d(gx_k,x_k)^2$ centered at the midpoint m between x_k and gx_k. If $h\in Q^{k-2}$, then the midpoint between hx_k and hgx_k is hm. Hence, since balls are convex, hm also lies in that ball centered at m. So $r_{k-2}^2\leq(r_k+\varepsilon)^2-\frac{1}{4}d(gx_k,x_k)^2$, and thus $\frac{1}{4}L(Q)^2\leq(r_k+\varepsilon)^2-r_{k-2}^2$. As ε was arbitrary, we get $\frac{1}{4}L(Q)^2\leq {r_k}^2-r_{k-2}^2$ and summing over even k, $\frac{n}{4}L(Q)^2\leq r_{2n}^2$ for all n, hence (1). But by Lemma 2.1 (b), $\Lambda_k(Q^{2n})\geq c^2E_k(Q^n)^2$, hence $R_k(Q)\geq\Lambda_k(Q^{2n})^{\frac{1}{2n}}\geq c^{\frac{1}{n}}E_k(Q)^{\sqrt{\frac{n-1}{8n^2}}}$. Optimizing in n, we obtain the desired bound. □

3. Height, arithmetic spectral radius and minimal height

For any rational prime p let us fix an algebraic closure $\overline{\mathbb{Q}_p}$ of the field of p-adic numbers $\mathbb{Q}_p$. We take the standard normalization of the absolute value on $\mathbb{Q}_p$ (i.e. $|p|_p=\frac{1}{p}$). It admits a unique extension to $\overline{\mathbb{Q}_p}$, which we denote by $|\cdot|_p$. Let $\overline{\mathbb{Q}}$ be the field of all algebraic numbers and K a number field. Let V_K be the set of equivalence classes of valuations on K. For $v\in V_K$ let K_v be the corresponding completion. For each $v\in V_K$, K_v is a finite extension of $\mathbb{Q}_p$ for some prime p. We normalize the absolute value on K_v to be the unique one which extends the standard absolute value on $\mathbb{Q}_p$. Namely $|x|_v=|N_{K_v|\mathbb{Q}_p}(x)|_p^{\frac{1}{n_v}}$ where $n_v=[K_v:\mathbb{Q}_p]$. Equivalently K_v has n_v different embeddings in $\overline{\mathbb{Q}_p}$ and each of them gives rise to the same absolute value on K_v. We identify $\overline{K_v}$, the algebraic closure of K_v with $\overline{\mathbb{Q}_p}$. Let V_f be the set of finite places and V_∞ the set of infinite places.

Let F be a finite subset in $M_2(K)$. For $v\in V_K$, in order not to surcharge notation, we will use the subscript v instead of K_v in the quantities $E_v(F)=E_{K_v}(F)$, $\Lambda_v(F)=\Lambda_{K_v}(F)$, etc.

Recall that if $x\in K$ then its (Weil-) height is by definition (see e.g. [3]) the following quantity

$$h(x)=\frac{1}{[K:\mathbb{Q}]}\sum_{v\in V_K}n_v\log^+|x|_v$$

It is well defined (i.e. independent of the choice of $K\ni x$). Let us similarly define the height of a matrix $f\in M_2(K)$ by

$$h(f)=\frac{1}{[K:\mathbb{Q}]}\sum_{v\in V_K}n_v\log^+||f||_v$$

and the height of a finite set F of matrices in $M_2(K)$ by

$$h(F)=\frac{1}{[K:\mathbb{Q}]}\sum_{v\in V_K}n_v\log^+||F||_v$$

where $n_v=[K_v:\mathbb{Q}_v]$. We also define the *minimal height* of F as:

$$e(F) = \frac{1}{[K:\mathbb{Q}]} \sum_{v \in V_K} n_v \log^+ E_v(F) \tag{2}$$

and the *arithmetic spectral radius* (or normalized height) of F as:

$$\widehat{h}(F) = \frac{1}{[K:\mathbb{Q}]} \sum_{v \in V_K} n_v \log^+ R_v(F)$$

For any height h, we also set $h = h_\infty + h_f$, where h_∞ is the infinite part of h (i.e. the part of the sum over the infinite places of K) and h_f is the finite part of h (i.e. the part of the sum over the finite places of K).

Note that these heights are well defined independently of the number field K such that $F \subset M_2(K)$. The above terminology is justified by the following facts:

Proposition 3.1. (Basic properties of heights I) For any finite set F in $M_2(\overline{\mathbb{Q}})$, we have:

(1) $\widehat{h}(F) = \lim_{n\to+\infty} \frac{1}{n} h(F^n) = \inf_{n \in \mathbb{N}} \frac{1}{n} h(F^n)$,

(2) $e_f(F) = \widehat{h}_f(F)$ and $e(F) + \log c \le \widehat{h}(F) \le e(F)$ where c is the constant in Lemma 2.1 (b),

(3) $\widehat{h}(F^n) = n \cdot \widehat{h}(F)$ and $\widehat{h}(F \cup \{Id\}) = \widehat{h}(F)$.

Proof. This follows directly from Proposition 2.5. □

We also record the following simple observations:

Proposition 3.2. (Basic properties of heights II) We have, for a finite set F in $M_2(\overline{\mathbb{Q}})$,

(1) $e(xFx^{-1}) = e(F)$ if $x \in GL_2(\overline{\mathbb{Q}})$.

(2) $e(F^n) \le n \cdot e(F)$,

(3) If λ is an eigenvalue of an element of F, then $h(\lambda) \le \widehat{h}(F) \le e(F)$,

Proof. This is clear. □

We can also compare $e(F)$ and $\widehat{h}(F)$ when $\widehat{h}(F)$ is small:

Proposition 3.3. (Basic properties of heights III) Let c_1 be the constant from Lemma 2.6, then

$$\widehat{h}_\infty(F) \ge \frac{c_1}{4} \cdot e_\infty(F) \cdot \min\{1, e_\infty(F)\}$$

for any finite subset F in $SL_2(\overline{\mathbb{Q}})$. In particular $e(F)$ is small as soon as $\widehat{h}(F)$ is small.

Proof. From Lemma 2.6, $\widehat{h}_v(F) \ge c_1 \cdot e_v(F) \cdot \min\{1, e_v(F)\}$ for every $v \in V_K$. We may write $e_\infty(F) = \alpha e^+(F) + (1-\alpha)e^-(F)$ where e^+ is the average of the e_v greater than 1 and e^- the average of the e_v smaller than 1. Applying Cauchy-Schwartz, we have $\widehat{h}_v(F) \geq c_1 \cdot (\alpha e^+ + (1-\alpha)(e^-)^2)$. If $\alpha e^+(F) \geq \frac{1}{2}e_\infty(F)$, then $\widehat{h}_v(F) \geq \frac{c_1}{2}e_\infty(F)$, and otherwise $(1-\alpha)e^- \geq \frac{e_\infty}{2}$, hence $\widehat{h}_v(F) \geq (1-\alpha)(e^-)^2 \geq \frac{1}{4}e_\infty^2$. At any case $\widehat{h}_\infty(F) \geq \frac{c_1}{4} \cdot e_\infty(F) \cdot \min\{1, e_\infty(F)\}$. □

4. Height Gap Theorem

In this section, we prove Theorem 1.2 from the Introduction. First observe that according to Propositions 3.3 and 3.1 (2), $\widehat{h}(F)$ is small if and only if $e(F)$ is small. So we may as well replace $\widehat{h}(F)$ by $e(F)$ in Theorem 1.2. We now assume that $F = \{Id, A, B\}$, with A semisimple. The general case follows from this as we will show in Lemma 4.7. Since $e(F)$ is invariant under conjugation by any element in $GL_2(\overline{\mathbb{Q}})$, we may assume that A is diagonal, i.e.

$$A = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix}, \ B = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{3}$$

The main part of the argument consists in the following proposition:

Proposition 4.1. (small normalised height implies small height of matrix coordinates) For every $\beta > 0$ there exists d_0,$\eta > 0$ such that, if $F = \{A, B\}$ are as in (3) and if $e(F) \leq \eta$ and $\deg(\lambda) \geq d_0$, then

$$\max\{h(ad), h(bc)\} \leq \beta$$

In order to prove this statement, we are first going to give local estimates at each place v, then use Bilu's equidistribution theorem to show that when these estimates are put together, the error terms give only a negligible contribution to the height.

Let K be the number field generated by the coefficients of A and B. Let $v \in V_K$ be a place of K. We set $s_v = \log E_v(F)$ and $\delta = \lambda - \lambda^{-1}$. We first show the following local estimate :

Lemma 4.2. (Local estimates) For each $v \in V_K$ we have

$$\max\{|a|_v, |d|_v, \sqrt{|bc|_v}\} \leq C_v e^{2s_v} \max\{1, |\delta^{-1}|_v\},$$

where C_v is a constant equal to 1 if v is a finite place and equal to a number $C_\infty > 1$ if v is infinite. Moreover there are absolute constants $\varepsilon_0 > 0$ and $C_0 > 0$ such that if v is infinite and $s_v \leq \varepsilon_0$, then

$$\max\{|ad|_v, |bc|_v\} \leq 1 + C_0(\sqrt{s_v} + \frac{\sqrt{s_v}}{|\delta|_v} + \frac{s_v}{|\delta|_v^2}).$$

Proof. In order not to overburden notation in this proof we set s_v to be some number arbitrarily close but strictly bigger than $\log E_v(F)$ and we can let it tend to $\log E_v(F)$ at the end. If v is infinite, then $\overline{\mathbb{Q}}_v = \mathbb{C}$ and $SL_2(\mathbb{C}) = \mathbf{K}AN$ where $\mathbf{K} = SU_2(\mathbb{C})$, A is the subgroup of diagonal matrices with real positive entries, and N is the subgroup of unipotent complex upper triangular matrices. As K leaves the norm invariant, there must exist a matrix $P \in AN$ such that $\max\{\|PAP^{-1}\|, \|PBP^{-1}\|\} \leq e^{s_v}$. Since $P \in AN$, we may write $P = \begin{pmatrix} t & y \\ 0 & t^{-1} \end{pmatrix}$ with $t > 0$ and $y \in \mathbb{C}$. Then we have, setting $\delta = \lambda - \lambda^{-1}$,

$$PAP^{-1} = \begin{pmatrix} \lambda & ty\delta \\ 0 & \lambda^{-1} \end{pmatrix}, \; PBP^{-1} = \begin{pmatrix} a + cyt^{-1} & bt^2 + dyt - ayt - cy^2 \\ t^{-2}c & -yct^{-1} + d \end{pmatrix}. \tag{4}$$

If v is finite and K_v is the corresponding completion, with ring of integers $\mathcal{O}_v$ and uniformizer π, we have $SL_2(K_v) = \mathbf{K}_v A_v N_v$ where $\mathbf{K}_v = SL_2(\mathcal{O}_v)$, $A_v = \{diag(\pi^n, \pi^{-n}), n \in \mathbb{Z}\}$ and N_v is the subgroup of unipotent upper triangular matrices with coefficients in K_v. Hence we also get a $P \in A_v N_v$ satisfying (4) with $y \in K_v$ and $t = \pi^n$ for some $n \in \mathbb{Z}$.

We first assume that v is finite. Recall that the operator norm in $SL_2(K_v)$ is given by the maximum modulus of each matrix coefficient. Hence we must have $|t^{-2}c|_v \leq e^{s_v}$ and $|ty\delta|_v \leq e^{s_v}$. It follows that $|cyt^{-1}|_v \leq e^{2s_v}|\delta^{-1}|_v$ and hence $|a|_v \leq \max\{e^{s_v}, e^{2s_v}|\delta^{-1}|_v\}$. Similarly, $|d|_v \leq \max\{e^{s_v}, e^{2s_v}|\delta^{-1}|_v\}$. Hence $|ad|_v \leq \max\{e^{2s_v}, e^{4s_v}|\delta^{-1}|_v^2\}$. Moreover $ad - bc = 1$, hence $|bc|_v \leq \max\{1, |ad|_v\} \leq \max\{e^{2s_v}, e^{4s_v}|\delta^{-1}|_v^2\}$.

Now we assume that v is infinite. **Claim:** There is $u_0 > 0$ such that if $0 \leq u \leq u_0$ and $\|B\| \leq e^u$, then

$$\max\{|a - \overline{d}|, |b + \overline{c}|\} \leq 2\sqrt{u} \tag{5}$$

$$\max\{|a|^2 + |b|^2, |d|^2 + |c|^2\} \leq 1 + 6u + 8\sqrt{u} \tag{6}$$

$$\max\{|a|, |b|, |c|, |d|\} \leq 1 + 3u + 4\sqrt{u} \leq 1 + 5\sqrt{u} \tag{7}$$

To prove this recall that the operator norm in $SL_2(\mathbb{C})$ satisfies $tr(B^*B) = |a|^2 + |b|^2 + |c|^2 + |d|^2 = \|B\|^2 + \|B\|^{-2}$. Hence $|a|^2 + ... + |d|^2 \leq 1 + e^{2u}$, hence ≤ 4 if u is small enough (say $u \leq .5$). On the other hand, for small u, $|a - \overline{d}|^2 + |b + \overline{c}|^2 = |a|^2 + ... + |d|^2 - 2 \leq e^{2u} - 1 \leq 4u$. Hence (5). Now $|d| \leq |a| + 2\sqrt{u}$ and since $|a|, |b| \leq 2$, we get $|d|^2 \leq |a|^2 + 4u + 8\sqrt{u}$ and vice versa and similarly for b and c. Hence (6) and (7) and the claim is proved.

Let now $\varepsilon > 0$ and assume that $s_v \leq \varepsilon$. From (4) we get $|\lambda|^2 + |\lambda^{-1}|^2 + |ty\delta|^2 \leq 1 + e^{2\varepsilon}$ hence $|ty\delta|^2 \leq e^{2\varepsilon} - 1 \leq 4\varepsilon$ if ε is small enough. So $|ty\delta| \leq 2\sqrt{\varepsilon}$. Now since $\|PBP^{-1}\| \leq e^{\varepsilon}$, we have $|t^{-2}c| \leq 2$ as soon as $\varepsilon \leq \frac{1}{2}$. Hence $|yct^{-1}| \leq \frac{4\sqrt{\varepsilon}}{|\delta|}$ and $\max\{|a|, |d|\} \leq 1 + 5\sqrt{\varepsilon} + \frac{4\sqrt{\varepsilon}}{|\delta|}$. Finally for some absolute constant $C > 0$ $|ad| \leq 1 + C(\sqrt{\varepsilon} + \frac{\sqrt{\varepsilon}}{|\delta|} + \frac{\varepsilon}{|\delta|^2})$.

On the other hand, $|cy^2| = |t^{-2}c(ty)^2| \leq \frac{8\varepsilon}{|\delta|^2}$ and $|d - a||yt| \leq \frac{24\sqrt{\varepsilon}}{|\delta|} + \frac{16\varepsilon}{|\delta|^2}$. Also by (5), $|bt^2 + (d - a)yt - cy^2 + t^{-2}\overline{c}| \leq 2\sqrt{\varepsilon}$, and $|bc + |t^{-2}c|^2| \leq 2|bt^2 + t^{-2}\overline{c}| \leq$

$4\sqrt{\varepsilon}+\frac{48\sqrt{\varepsilon}}{|\delta|}+\frac{48\varepsilon}{|\delta|^2}$, and by (6), $|t^{-2}c|^2 \leq 1+14\sqrt{\varepsilon}$, hence up to enlarging the absolute constant C, we also have $|bc| \leq 1 + C(\sqrt{\varepsilon} + \frac{\sqrt{\varepsilon}}{|\delta|} + \frac{\varepsilon}{|\delta|^2})$.

Without the assumption that s_v is small, we can make a coarser estimate: $|t^{-2}c|^2 \leq 1 + e^{2s_v}$, $|ty\delta|^2 \leq 1 + e^{2s_v}$, hence $|cyt^{-1}| \leq \frac{1+e^{2s_v}}{|\delta|}$ and $\max\{|a|, |d|\} \leq \frac{1+e^{2s_v}}{|\delta|} + \sqrt{1 + e^{2s_v}} \leq 2e^{2s_v} \max\{1, \frac{1}{|\delta|}\}$ and $|ad| \leq 4e^{4s_v} \max\{1, \frac{1}{|\delta|^2}\}$. Similarly, we compute $|bc| \leq 20e^{4s_v} \max\{1, \frac{1}{|\delta|^2}\}$.

□

We now put together the local information obtained above to bound the heights. Let $n = [K : \mathbb{Q}]$ and V_f and V_∞ the set of finite and infinite places of K. Set ε_0, C_0 and C_∞ the constants obtained in the previous lemma. For $A > 0$ and $x \in \overline{\mathbb{Q}}$, we set

$$h_\infty^A(x) = \frac{1}{[K : \mathbb{Q}]} \sum_{v \in V_\infty, |x|_v \geq A} n_v \cdot \log^+ |x|_v \tag{8}$$

where the sum is limited to those $v \in V_\infty$, for which $|x|_v \geq A$. We have:

Lemma 4.3. For some constant C_2 $(2 \leq C_2 \leq 2 + (2 \log C_\infty + 4)/\log 2)$, we have for all $\varepsilon_1 \in (0, \frac{1}{2})$ and all $\varepsilon \leq \min\{\varepsilon_0, \varepsilon_1^2\}$

$$\max\{h(ad), h(bc)\} \leq C_{\varepsilon,\varepsilon_1} e(F) + 6C_0 \frac{\sqrt{\varepsilon}}{\varepsilon_1} + 2h_f(\delta^{-1}) + C_2 \cdot h_\infty^{\varepsilon_1^{-1}}(\delta^{-1}) \tag{9}$$

where $C_{\varepsilon,\varepsilon_1} = \left(12 + \frac{2\log C_\infty}{\varepsilon} + \frac{2|\log \varepsilon_1|}{\varepsilon}\right)$ and $\delta = \lambda - \lambda^{-1}$.

Proof. If $v \in V_\infty$ and $s_v \geq \varepsilon$, then according to Lemma 4.2 $\log^+ |ad|_v \leq 2\log C_\infty + 4s_v + 2\log^+ |\delta^{-1}|_v$ hence

$$\frac{1}{n} \sum_{v \in V_\infty, s_v \geq \varepsilon} n_v \cdot \log^+ |ad|_v \leq \left(4 + \frac{2\log C_\infty}{\varepsilon}\right) \frac{1}{n} \sum_{v \in V_\infty, s_v \geq \varepsilon} n_v s_v + \frac{2}{n} \sum_{v \in V_\infty, s_v \geq \varepsilon} n_v \cdot \log^+ |\delta^{-1}|_v$$

Fix $\varepsilon_1 < \frac{1}{2}$. On the other hand, if $s_v \leq \varepsilon \leq \min\{\varepsilon_0, \varepsilon_1^2\}$ and $|\delta|_v \geq \varepsilon_1$ then $\log^+ |ad|_v \leq C_0(\sqrt{s_v} + \frac{\sqrt{s_v}}{|\delta|} + \frac{s_v}{|\delta|^2}) \leq 3C_0 \frac{\sqrt{\varepsilon}}{\varepsilon_1}$ and, as $n_v \leq 2$,

$$\frac{1}{n} \sum_{v \in V_\infty, s_v \leq \varepsilon, |\delta|_v \geq \varepsilon_1} n_v \cdot \log^+ |ad|_v \leq 6C_0 \frac{\sqrt{\varepsilon}}{\varepsilon_1}$$

While if $s_v < \varepsilon$ and $|\delta|_v \leq \varepsilon_1 \leq \frac{1}{2}$, then $\log^+ |ad|_v \leq C_2 \log^+ |\delta^{-1}|_v$ for some absolute constant C_2, $(2 \leq C_2 \leq 2 + (2\log C_\infty + 4)/\log 2)$, hence

$$\frac{1}{n} \sum_{v \in V_\infty, s_v < \varepsilon, |\delta|_v \leq \varepsilon_1} n_v \cdot \log^+ |ad|_v \leq \frac{1}{n} \sum_{v \in V_\infty, s_v < \varepsilon, |\delta|_v \leq \varepsilon_1} C_2 n_v \cdot \log^+ |\delta^{-1}|_v$$

When $v \in V_f$, from Lemma 4.2, we get

$$\sum_{v \in V_f} n_v \cdot \log^+ |ad|_v \leq \sum_{v \in V_f} 4 n_v s_v + \sum_{v \in V_f} 2 n_v \cdot \log^+ |\delta^{-1}|_v$$

But

$$\frac{2}{n} \sum_{v \in V_\infty, s_v \geq \varepsilon, |\delta|_v \geq \varepsilon_1} n_v \cdot \log^+ |\delta^{-1}|_v \leq \frac{2|\log \varepsilon_1|}{\varepsilon} \frac{1}{n} \sum_{v \in V_\infty, s_v \geq \varepsilon} n_v s_v$$

Putting together the above estimates, we indeed obtain (9) for ad. The same computation works for bc. □

It is now time to recall the following result (see also [12] and [25]):

Theorem 4.4. (**Bilu's equidistribution of small points**, [2]) Suppose $(\lambda_n)_{n \geq 1}$ is a sequence of algebraic numbers (i.e. in $\overline{\mathbb{Q}}$) such that $h(\lambda_n) \to 0$ and $\deg(\lambda_n) \to +\infty$ as $n \to +\infty$. Let $\mathcal{O}(\lambda_n)$ be the Galois orbit of λ_n in $\overline{\mathbb{Q}}$. Then we have the following weak-$*$ convergence of probability measures on $\mathbb{C}$,

$$\frac{1}{\#\mathcal{O}(\lambda_n)} \sum_{x \in \mathcal{O}(\lambda_n)} \delta_x \underset{n \to +\infty}{\to} d\theta \tag{10}$$

where $d\theta$ is the normalized Lebesgue measure on the unit circle $\{z \in \mathbb{C}, |z| = 1\}$.

We now draw two consequences of this equidistribution statement :

Lemma 4.5. (bounding errors terms via Bilu's theorem I) For every $\alpha > 0$ there is d_1,$\eta_1 > 0$ and $\varepsilon_1 > 0$ with the following property. If $\lambda \in \overline{\mathbb{Q}}$ is such that $h(\lambda) \leq \eta_1$, $\deg(\lambda) \geq d_1$ then

$$h_\infty^{\varepsilon_1^{-1}}(\frac{1}{1-\lambda}) \leq \alpha$$

where $h_\infty^{\varepsilon_1^{-1}}$ was defined in (8).

Proof. Let $P \in \mathbb{Z}[X]$ be the minimal polynomial of λ, i.e. $P(X) = \sum_{0 \leq i \leq n} a_i X^i = a_n \prod_{x \in \mathcal{O}(\lambda)} (X - x)$. As $P(1) \in \mathbb{Z} \backslash \{0\}$, $\log |P(1)| = \log |a_n| + \sum_{x \in \mathcal{O}(\lambda)} \log |1 - x| \geq 0$. So

$$\sum_{|1-x| \leq \varepsilon_1} \log \frac{1}{|1-x|} \leq \sum_{|1-x| > \varepsilon_1} \log |1-x| + \log |a_n|$$

Recall (see [22] III.1.) that $h(\lambda) = \frac{1}{n} \left(\sum_{x \in \mathcal{O}(\lambda)} \log^+ |x| + \log |a_n| \right)$. Hence

$$\frac{1}{n} \sum_{|1-x| \leq \varepsilon_1} \log \frac{1}{|1-x|} \leq h(\lambda) + \frac{1}{n} \sum_{|1-x| > \varepsilon_1} \log |1-x| \tag{11}$$

Consider the function $f_{\varepsilon_1}(z) = \mathbf{1}_{|z-1| > \varepsilon_1} \log |1 - z|$. It is locally bounded on $\mathbb{C}$. By Theorem 4.4, for every $\varepsilon_1 > 0$, there must exist $d_1, \eta_1 > 0$ such that, if $h(\lambda) \leq \eta_1$ and $\deg(\lambda) \geq d_1$, then $\left| \frac{1}{n} \sum_x f_{\varepsilon_1}(x) - \int_0^1 f_{\varepsilon_1}(e^{2\pi i \theta}) d\theta \right| \leq \frac{\alpha}{3}$. On the other hand we verify that $\theta \mapsto \log|1-e^{2\pi i\theta}|$ is in $L^1(0,1)$ and $\int_0^1 \log|1-e^{2\pi i\theta}|d\theta = 0$. Hence we can choose $\varepsilon_1 > 0$ small enough so that $\left|\int_0^1 f_{\varepsilon_1}(e^{2\pi i\theta})d\theta\right| \leq \frac{\alpha}{3}$. Combining these inequalities with (11) and choosing $\eta_1 \leq \frac{\alpha}{3}$, we get $h_\infty^{\varepsilon_1^{-1}}((1-\lambda)^{-1}) \leq \alpha$.

□

Combining this with Bilu's theorem, we get:

Lemma 4.6. (bounding errors terms via Bilu's theorem II) For every $\alpha > 0$ there exists $\eta_0 > 0$ and $A_1 > 0$ such that for any $\lambda \in \overline{\mathbb{Q}}$, if $h(\lambda) \leq \eta_0$ and $d = \deg(\lambda) > A_1$, then

$$h_f(\frac{1}{1-\lambda}) \leq 2\alpha$$

Proof. We apply the *product formula* to $\mu = 1-\lambda$, which takes the form $h(\mu) = h(\mu^{-1})$, hence $h_f(\mu^{-1}) = h_\infty(\mu) - h_\infty(\mu^{-1}) + h_f(\mu)$. But $h_f(\mu) = h_f(1-\lambda) \leq h_f(\lambda) \leq \eta_0$ and $h_\infty(\mu) - h_\infty(\mu^{-1}) = \frac{1}{[K:\mathbb{Q}]}\sum_{v\in V_\infty} n_v \cdot \log|\mu|_v$. Lemma 4.5 shows that the convergence (10) in Bilu's theorem not only holds for compactly supported functions on $\mathbb{C}$, but also for functions with logarithmic singularities at 1. In particular it holds for the function $f(z) = \log|1-z|$, which is exactly what we need, since $\int_0^1 f(e^{2\pi i\theta})d\theta = 0$. Hence $\frac{1}{[K:\mathbb{Q}]}\sum_{v\in V_\infty} n_v \cdot \log|\mu|_v$ becomes small. We are done. □

Proof of Proposition 4.1. Since $h_f(\frac{1}{\lambda-\lambda^{-1}}) \leq h_f(\lambda) + h_f(\frac{1}{1-\lambda^2})$ and similarly $h^A_\infty(\frac{1}{\lambda-\lambda^{-1}}) \leq h^A_\infty(\lambda) + h^A_\infty(\frac{1}{1-\lambda^2})$, it follows from the last two lemmas that we can find $\varepsilon_1 > 0$, $\eta > 0$ and $d_0 \in \mathbb{N}$ so that $2h_f(\delta^{-1}) + C_2 \cdot h_\infty^{\varepsilon_1^{-1}}(\delta^{-1}) \leq \frac{\beta}{3}$ as soon as $h(\lambda) \leq e(F) \leq \eta$ and $\deg(\lambda) \geq d_0$. Then choose ε so the $2C_1\frac{\sqrt{\varepsilon}}{\varepsilon_1} \leq \frac{\beta}{3}$ and finally take η even smaller so that $C_{\varepsilon,\varepsilon_1}\eta \leq \frac{\beta}{3}$. Now apply Lemma 4.3 and we are done. □

End of the proof of Theorem 1.2: The following lemma allows us, when proving Theorem 1.2, to assume without loss of generality that $F = \{1, A, B\}$, where A and B are two semisimple elements in $SL_2(\overline{\mathbb{Q}})$ that do not satisfy some prescribed finite set of algebraic relations. More precisely:

Lemma 4.7. For every $d_1 \in \mathbb{N}$, there exists $N(d_1) \in \mathbb{N}$ with the following property. Let F be a finite subset of $SL_2(\overline{\mathbb{Q}})$ containing 1 and generating a non-virtually solvable subgroup, then there exists $A, B \in F^{N(d_1)}$ such that A and B are semisimple, generate a non-virtually solvable subgroup of SL_2, A is not of order at most d_1, and $bc \notin \{0, -1, e^{\frac{2i\pi}{3}}, e^{\frac{4i\pi}{3}}\}$ after we conjugate A and B in the form (3).

Proof. This is a direct application of Lemma 4.8 below applied to $\Sigma = F \times F$ in $SL_2 \times SL_2 \leq GL_4$ with $X = X_1 \cup X_2 \cup X_3 \cup X_4$ where $X_1 = \{(A,B),\ A$ or B has order at most $d_1\}$, $X_2 = \{(A,B),\ tr(A)$ or $tr(B)$ is $2\}$, $X_3 = \{(A,B),$ A and B generate a virtually solvable subgroup$\}$ and X_4 the Zariski closure of

$\{(gAg^{-1}, gBg^{-1}),\ g \in SL_2,\ A$ diagonal, $bc \in \{0, -1, e^{\frac{2i\pi}{3}}, e^{\frac{4i\pi}{3}}\}\}$. For dimension reasons X_4 is a proper subvariety of $SL_2 \times SL_2$ and Propositions 9.2 and 9.1 show that so is X_3. □

Lemma 4.8. (**Eskin-Mozes-Oh "Escape from subvarieties"**, see [13] and [9]) Let K be a field, $d \in \mathbb{N}$. For every $m \in \mathbb{N}$, there is $N \in \mathbb{N}$ such that if X a K–algebraic subvariety of $GL_d(K)$ such that the sum of the degrees of the geometrically irreducible components of X is at most m, then for any subset $\Sigma \subset GL_d(K)$ containing Id and generating a subgroup which is not contained in $X(K)$, we have $\Sigma^N \nsubseteq X(K)$.

Observe that for every $d_0 \in \mathbb{N}$ there is $\eta_0 > 0$ and $d_1 > 0$ such that if $h(\lambda) < \eta_0$ and λ is not a root of 1 of order at most d_1, then $\deg(\lambda) \geq d_0$. However, recall the following well known result (which is also a straightforward corollary of Theorem 4.4),

Theorem 4.9. (**Zhang's theorem** [32]) There exists an absolute constant $\alpha_0 > 0$ such that for any $x \in \overline{\mathbb{Q}}$, we have

$$h(x) + h(1 + x) > \alpha_0$$

unless $x \in \{0, -1, e^{\frac{2i\pi}{3}}, e^{\frac{4i\pi}{3}}\}$.

Let $\beta = \frac{\alpha_0}{2}$ where α_0 is given by Theorem 4.9. Proposition 4.1 yields $d_0 > 0$ and $\eta = \eta(\frac{\alpha_0}{2}) > 0$ such that $\max\{h(ad), h(bc)\} \leq \beta$ as soon as $e(\{Id, A, B\}) \leq \eta$ and $\deg(\lambda) \geq d_0$. By Lemma 4.7, if we have some nice $A, B \in F^{N(d_1)}$. If $e(F) \leq \frac{\min\{\eta(\frac{\alpha_0}{2}), \eta_0\}}{N(d_1)}$, then $e(\{Id, A, B\}) \leq \min\{\eta, \eta_0\}$ and λ is not a root of 1 of order at most d_1. Hence $\deg(\lambda) \geq d_0$ and by Proposition 4.1, $h(ad) + h(bc) \leq 2\beta = \alpha_0$. Then according to Theorem 4.9, $bc \in \{0, -1, e^{\frac{2i\pi}{3}}, e^{\frac{4i\pi}{3}}\}$, which contradicts our choice of A, B. So $\frac{\min\{\eta(\frac{\alpha_0}{2}), \eta_0\}}{N(d_1)} > 0$ is the desired gap.

This ends the proof of Theorem 1.2.

Finally observe that Theorem 1.2 combined with Lemma 3.1 to 3.3 implies :

Proposition 4.10. There exists a constant $c_0 > 0$ such that if F is any finite subset of $SL_2(\overline{\mathbb{Q}})$ generating a non-virtually solvable subgroup, then

$$e(F) \geq \widehat{h}(F) \geq c_0 \cdot e(F)$$

5. Simultaneous quasi-symmetrization over $\overline{\mathbb{Q}}$

Here we are going to use our previous height estimates once again to show the following proposition. Observe that the minimal height $e(F)$ coincides with the infimum of $h(gFg^{-1})$ over all adelic points $g = (g_v)_v$. The lemma we are about to state essentially means that this infimum is attained (up to additive and multiplicative constants) with a conjugating matrix g lying already in $SL_2(\overline{\mathbb{Q}})$ as opposed to $SL_2(\mathbb{A})$ (the adelic group). When a matrix with real entries is symmetric, then its norm coincides with the modulus of its maximal eigenvalue. Thus the lemma amounts to conjugating F simultaneously (i.e. by a single $g \in SL_2(\overline{\mathbb{Q}})$) in a "quasi-symmetric" position.

Proposition 5.1. (Simultaneous quasi-symmetrization) There is an absolute constant $C > 0$ such that if F is a finite subset of $SL_2(\overline{\mathbb{Q}})$ generating a non-virtually solvable subgroup, then there is an element $g \in SL_2(\overline{\mathbb{Q}})$ such that

$$h(gFg^{-1}) \leq C \cdot e(F) + C$$

Proof. As we may replace F by a bounded power of it, Lemma 4.7 above allows us to assume that F contains a semisimple element. Let $F = \{Id, A, B_1, ..., B_k\}$ with A semisimple. Conjugating by some $g \in SL_2(\overline{\mathbb{Q}})$, we may assume that A is in diagonal form and we write each B_i in the form (3) with entries a_i, b_i, c_i, d_i. Changing F into F^2 if necessary, we may assume that both b_1 and c_1 are not zero (otherwise F would be contained in the group of upper or lower triangular matrices). We may further conjugate F by the diagonal matrix $diag(t, t^{-1})$, where $t \in \overline{\mathbb{Q}}$ is a root of $t^4 = c_1/b_1$, so as to ensure $b_1 = c_1$. Then $h(B_1) \leq h(a_1) + h(d_1) + 2h(b_1) + \log 2$. On the other hand, since $a_1d_1 - b_1c_1 = 1$, we have $b_1^2 = a_1d_1 - 1$ and $2h(b_1) = h(b_1^2) \leq h(a_1d_1) + \log 2 \leq 2e(\{A, B\}) + \log 2C_\infty$. On the other hand, by Lemma 4.2 applied to $\{A, B_i\}$ we have $\max\{|a_i|_v, |d_i|_v\} \leq C_v e^{2s_v} \max\{1, |\delta^{-1}|_v\}$, for every place v, where $\delta = \lambda - \lambda^{-1}$ and $s_v = s_v(\{A, B_i\}) = \log E_v(\{A, B_i\})$. Applying Lemma 4.2 to $\{A, B_1B_i\}$ we get $\max\{|(B_1B_i)_{11}|_v, |(B_1B_i)_{22}|_v\} \leq C_v e^{2s_v} \max\{1, |\delta^{-1}|_v\}$ with $s_v = s_v(\{A, B_1B_i\}) = \log E_v(\{A, B_1B_i\})$. We compute the matrix entry $(B_1B_i)_{11} = a_1a_i + b_1c_i$. We get

$$|c_i|_v = |[(B_1B_i)_{11} - a_1a_i]b_1^{-1}|_v \leq C_v e^{2s_v} \max\{1, |\delta^{-1}|_v\} \max\{1, |b_1^{-1}|_v\}$$

Similarly for $|b_i|_v$. Hence,

$$\begin{aligned} ||F||_v &\leq C_v \max_{i=1,\ldots,k} \{|\lambda|_v, |\lambda^{-1}|_v, |a_i|_v, |d_i|_v, |b_i|_v, |c_i|_v\} \\ &\leq C_v \max_{i=1,\ldots,k} E_v(\{A, B_1, B_1B_i\})^2 \cdot \max\{1, |\delta^{-1}|_v\} \max\{1, |b_1^{-1}|_v\} \end{aligned}$$

In particular this means that

$$\begin{aligned} h(F) &\leq 2\log C_\infty + 2e(F^2) + h(\delta) + h(b_1) \\ &\leq 7e(F) + 4\log 2C_\infty \end{aligned}$$

So we are done. □

Corollary 5.2. There exists a constant $C_{qs} > 0$ such that if F is as in the Proposition, then there is an element $g \in SL_2(\overline{\mathbb{Q}})$ such that

$$h(gFg^{-1}) \leq C_{qs} \cdot e(F)$$

Proof. It is clear from the combination of the previous proposition and Theorem 1.2. □

6. Ping-Pong

Here we state and prove a ping-pong criterion, which gives a sufficient condition on the finite set F for it, or a bounded power of it, to contain two free generators of a free subgroup. Let $k_1, k_2, k_3 \in \mathbb{N}$ be three positive integers and let k be a local field of characteristic zero with its standard absolute value. We set $C_k = 2$ if k is Archimedean and $C_k = 1$ if k is non Archimedean. Let $F \subset SL_2(k)$ be a finite set containing 1 such that $\Lambda_k(F^{k_1}) > C_k||F||_k$ (see Section 2 for notation, it is important to require a strict inequality here when k is non Archimedean). Let $A \in F^{k_1}$ be such that $\Lambda_k(A) = \Lambda_k(F^{k_1})$. Then of course A is semisimple and admits two distinct eigenvectors v^+ and v^- in k_q^2 where k_q is either k or some quadratic extension of k. Since we may always replace k by k_q, there is no loss of generality in assuming that v^+ and v^- lie in k^2. Let d_k be the canonical (Fubini-Study) projective distance on $\mathbb{P}^1(k)$, namely $d_k(u,v) = \frac{||u\wedge v||_k}{||u||_k||v||_k}$.

Lemma 6.1. (geometric conditions for ping-pong) Assume that there is $B \in F^{k_2}$ such that $d_k(Bv^\varepsilon, v^{\varepsilon'}) \geq ||F||_k^{-k_3}$, and $d_k(v^\varepsilon, v^{\varepsilon'}) \geq ||F||_k^{-k_3}$ for each $\varepsilon, \varepsilon' \in \{\pm\}$. Then A^l and BA^lB^{-1} play ping-pong on $\mathbb{P}^1(k)$ and generate a free subgroup of $SL_2(k)$ as soon as $l \geq (k_2+1)(k_3+1)$.

Proof. Note that $\forall u, v \in \mathbb{P}^1(k)$ we have $d_k(Bu, Bv) \leq ||B||^2 d_k(u,v)$ for $B \in SL_2(k)$. Note also that without loss of generality, we may assume that $||v^+||_k = ||v^-||_k = 1$. Let λ, λ^{-1} be the eigenvalues of A, where we have chosen $|\lambda|_k \geq 1$. By the assumption on A, $|\lambda| > C_k||F||_k \geq 1$. Since the roles of v^+ and v^- are interchangeable, we may assume that v^+ corresponds to λ and v^- to λ^{-1}. Let $P \in GL_2(k)$ be defined by $Pe_1 = v^+$ and $Pe_2 = v^-$. Note that $|\det P| = ||v^+ \wedge v^-|| = d_k(v^+, v^-)$. Also $||P|| = 1$ if k is non Archimedean, and $||P||^2 \leq 2$ if k is Archimedean, so in general $||P||^2 \leq C_k$. Moreover $||P^{-1}|| = ||P||/|\det P|_k \leq C_k||F||^{k_3}$. Set $A' = P^{-1}AP$, $B' = P^{-1}BP$, $F' = P^{-1}FP$, then $A' = diag(\lambda, \lambda^{-1})$.

For $u, v \in \mathbb{P}^1(k)$, $d_k(Pu, Pv) = \frac{||Pu\wedge Pv||}{||Pu||\cdot||Pv||} \leq |\det P|||P^{-1}||^2 d_k(u,v) \leq \frac{C_k\cdot d_k(u,v)}{|\det P|}$. Hence for $i, j \in \{1,2\}$,

$$d_k(B'e_i, e_j) \geq \frac{1}{C_k} d_k(v^+, v^-) d_k(BPe_i, Pe_j) \geq \frac{1}{C_k}\frac{1}{||F||^{2k_3}}. \tag{12}$$

Observe also that $||F'|| \leq ||F|| \cdot ||P||^2/|\det P| \leq C_k||F||^{k_3+1}$.

Let $m \leq 2l$ be positive integers to be determined shortly below. Let $\mathcal{U}_A^+ = \{x \in \mathbb{P}^1(k),\, d_k(x, e_1) \leq |\lambda|^{-2l}\}$, $\mathcal{U}_A^- = \{x \in \mathbb{P}^1(k),\, d_k(x, e_2) \leq |\lambda|^{-2l}\}$, $\mathcal{U}_C^+ = \{x \in \mathbb{P}^1(k),\, d_k(x, B'e_1) \leq |\lambda|^{-m}\}$ and $\mathcal{U}_C^- = \{x \in \mathbb{P}^1(k),\, d_k(x, B'e_2) \leq |\lambda|^{-m}\}$. We need to show that these four sets are disjoint, and that A'^l maps $(\mathcal{U}_A^-)^c$ into $\mathcal{U}_A^+$, A'^{-l} maps $(\mathcal{U}_A^+)^c$ into $\mathcal{U}_A^-$, $C' = B'A'^lB'^{-1}$ maps $(\mathcal{U}_C^-)^c$ into $\mathcal{U}_C^+$ and C'^{-1} maps $(\mathcal{U}_C^+)^c$ into $\mathcal{U}_C^-$.

If for instance $\mathcal{U}_A^+ \cap \mathcal{U}_C^- \neq \emptyset$, then $d(B'e_i, e_j) \leq \frac{C_k}{|\lambda|^m}$ for some i, j, which in turn would contradict (12) since $|\lambda|^m > C_k^2 ||F||^{2k_3}$ as soon as $m \geq 2k_3$. The same holds in other situations as soon as $m \geq 2(k_3+1)$.

Now since A' is diagonal, A' maps $(\mathcal{U}_A^-)^c$ into $\mathcal{U}_A^+$, A'^{-l} maps $(\mathcal{U}_A^+)^c$ into $\mathcal{U}_A^-$. Finally let us check the last two conditions. If $x \in (\mathcal{U}_C^-)^c$, then $d_k(x, B'e_2) > |\lambda|^{-m}$ and $d_k(B'^{-1}x, e_2)||B'||^2 > |\lambda|^{-m}$. So $B'^{-1}x \in (\mathcal{U}_A^-)^c$ as long as $|\lambda|^{2l-m} \geq ||B'||^2$. Then $A'^l B'^{-1} x \in \mathcal{U}_A^+$ and $d_k(C'x, B'e_1) \leq ||B'||^2/|\lambda|^{2l} \leq |\lambda|^{-m}$. And similarly if $x \in (\mathcal{U}_C^+)^c$.

So the above works as soon as $m \geq 2(k_3+1)$ (so that $|\lambda|^m > C_k^2||F||^{2(k_3+1)}$) and $2l - m \geq 2k_2(k_3+1)$ (so that $|\lambda|^{2l-m} > C_k^{2k_2}||F||^{2k_2k_3+2k_2} \geq ||F'||^{2k_2} \geq ||B'||^2$). □

Remark 6.2. A similar ping-pong lemma holds with the ping pong players A^l and BA^lB (instead of BA^lB^{-1}) if we assume similar lower bounds on $d_k(B^\delta v^\varepsilon, v^{\varepsilon'})$ for $\delta \in \{0, \pm 1, \pm 2\}$ and $\varepsilon, \varepsilon' \in \{\pm\}$. This allows to find the ping pong players in some F^n, i.e. without having to take inverses of elements of F.

6.1. **Quasi-isometrically embedded free subgroup.** A free subgroup H generated by two free elements a and b in a group Γ with finite generating set F (assumed symmetric) is said to be C-quasi isometrically embedded if for all $h \in H$

$$\frac{1}{C} \cdot d_\Gamma(1,h) \leq d_H(1,h) \leq C \cdot d_\Gamma(1,h)$$

where d_Γ is the word metric in Γ associated to F and d_H the word metric in H corresponding to the generating set $\{a^{\pm 1}, b^{\pm 1}\}$. In the setting of Lemma 6.1 we have:

Lemma 6.3. (QI embedding of free subgroup) The two elements A^l and BA^lB^{-1} generate a free subgroup H which is C-quasi isometrically embedded in the group Γ with generating set F with $C = 2k_2 + k_1 l$. More precisely,

$$\frac{1}{C} \cdot d_\Gamma(1,h) \leq d_H(1,h) \leq 4 \cdot d_\Gamma(1,h)$$

Proof. The inequality on the left hand side is clear as $a := A^l$ and $b := BA^lB^{-1}$ both belong to F^C. To prove the inequality on the right hand side, observe that both a and b act on the complement of their repelling neighborhood by transformations that contract distances by a factor at least $\frac{1}{|\lambda|_k} \leq \frac{1}{||F||_k}$. This implies that any element h that can be written as $h = w(a,b)$ for some reduced word w of length $n = d_H(1,h)$ in the free group will act on some open subset of $\mathbb{P}^1(k)$ by contracting distances by a factor at least $\frac{1}{||F||_k^n}$, and in particular $Lip(h) \geq ||F||_k^n$, where $Lip(h)$ is the bi-Lipschitz constant of h acting on $\mathbb{P}^1(k)$, $Lip(h) = \sup\{\left(\frac{d(hx,hy)}{d(x,y)}\right)^{\pm 1} | x, y \in \mathbb{P}^1(k)\}$. On the other, one easily checks that for any $g \in SL_2(k)$, $Lip(g) \leq ||g||_k^4$, hence $Lip(F^n) \leq ||F||_k^{4n}$ for all n and hence $Lip(h) \leq ||F||_k^{4d_\Gamma(1,h)}$ which yields $d_H(1,h) \leq 4 \cdot d_\Gamma(1,h)$ as desired. □

7. Proof of Theorems 1.1 and 1.3 for $\mathrm{SL}_2(\mathbb{C})$

We first assume that F has coefficients in $\overline{\mathbb{Q}}$. We explain at the end of this section why this case implies the general case.

We are going to show that if F generates a non virtually solvable subgroup of $SL_2(K)$ for some number field K, then for at least one place $v \in V_K$ the conditions of the ping-pong lemma 6.1 are satisfied, with k_1, k_2 and k_3 bounded and independent of K. This will be done by finding an appropriate prime and a place above it where F will satisfy the requirements of Lemma 6.1.

Let F be a finite subset of $SL_2(\overline{\mathbb{Q}})$ which generates a non virtually solvable subgroup and contains 1. According to Lemma 4.7, as one may change F into a bounded power of itself if necessary, we may assume that F contains two semisimple elements which generate a non virtually solvable subgroup. Now, from Corollary 5.2, after possibly conjugating F inside $SL_2(\overline{\mathbb{Q}})$, we may assume that $h(F) \leq C_{qs} \cdot e(F)$, where $C_{qs} > 0$ is the universal constant given by Corollary 5.2.

The last important ingredient in the proof of Theorem 1.1 is the product formula on the projective line $\mathbb{P}^1(\overline{\mathbb{Q}})$ (see [3]), i.e. $\forall (u,v) \in \mathbb{P}^1(\overline{\mathbb{Q}})^2$

$$\prod_{v \in V_K} d_v(u,v)^{\frac{1}{[K:\mathbb{Q}]}} = \frac{1}{H(u) \cdot H(v)} \tag{13}$$

where $\log H(u) = h(u) = \frac{1}{[K:\mathbb{Q}]} \sum_{v \in V_K} n_v \log \max\{|u_1|_v, |u_2|_v\}$ if $(u_1, u_2) \in K^2$ represents $u \in \mathbb{P}^1(K)$. This formula is straightforward from the usual product formula and the definition of the standard distance $d_v(u,v) = \frac{||u \wedge v||_{K_v}}{||u||_{K_v} ||v||_{K_v}}$.

Lemma 7.1. (Height of F controls heights of eigenobjects) Let $A \in SL_2(\overline{\mathbb{Q}})$ and $v \in \mathbb{P}^1(\overline{\mathbb{Q}})$ an eigendirection of A, then $h(v) \leq 3h(A) + \log 2$.

Proof. Simply solve for v in $Av = \lambda v$ using Cramer's rule. □

Let us introduce some notation. Suppose $A \in SL_2(\overline{\mathbb{Q}})$ is semisimple with eigendirections v_A^+ and v_A^- in $\mathbb{P}^1(\overline{\mathbb{Q}})$ and suppose $B \in SL_2(\overline{\mathbb{Q}})$. Then, assuming A and B have coefficients in a number field K, we set for each place $v \in V_K$:

$$\delta_v^{+,-}(B; A) = \log \frac{1}{d_v(Bv_A^+, v_A^-)}$$

where d_v is the standard distance on $\mathbb{P}^1(K_v)$ and K_v is the completion of K at v. Note that as $d_v \leq 1$, we have $\delta_v^{+,-}(B;A) \geq 0$. If $d_v(Bv_A^+, v_A^-) = 0$ we set $\delta_v^{+,-}(B;A) = 0$. We define similarly $\delta_v^{+,+}(B;A)$, $\delta_v^{-,+}(B;A)$, and $\delta_v^{-,-}(B;A)$ in the obvious manner and we set

$$\delta_v(B;A) = \delta_v^{+,-}(B;A) + \delta_v^{+,+}(B;A) + \delta_v^{-,+}(B;A) + \delta_v^{-,-}(B;A)$$

For a finite subset F of $SL_2(\overline{\mathbb{Q}})$, we also define

$$\delta_v(F) = \sum \delta_v(Id; A) + \delta_v(B; A)$$

where the sum runs over all pairs $\{A, B\}$ of elements of F with A semisimple and B in *"nice position" with respect to* A, namely such that $Bv_A^+ \notin \{v_A^+, v_A^-\}$ and $Bv_A^- \notin \{v_A^+, v_A^-\}$. If this set of pairs is empty we set δ to be 0. However, in our case, it will be non empty if not for F itself then for a bounded power of it (see Lemma 7.3 below). We also define the corresponding global quantity:

$$\delta(B; A) = \frac{1}{[K : \mathbb{Q}]} \sum_{v \in V_K} n_v \cdot \delta_v(B; A)$$

and

$$\delta(F) = \frac{1}{[K : \mathbb{Q}]} \sum_{v \in V_K} n_v \cdot \delta_v(F)$$

Proposition 7.2. (Height of F controls adelic distance between eigenobjects) With the above notation, for every $B \in SL_2(\overline{\mathbb{Q}})$ in nice position with respect to a semisimple $A \in SL_2(\overline{\mathbb{Q}})$ (or for $B = Id$), we have

$$\delta(B; A) \leq 24h(A) + 4h(B) + 12 \log 2$$

In particular for any finite subset F in $SL_2(\overline{\mathbb{Q}})$

$$\delta(F) \leq 12|F|^2(3h(F) + \log 2)$$

Proof. From the product formula (13) above we have $\delta^{+,-}(B; A) = h(Bv_A^+) + h(v_A^-)$. On the other hand we easily compute $h(Bv_A^+) \leq h(B) + h(v_A^+) + \log 2$. From Lemma 7.1, we get $\delta^{+,-}(B; A) \leq h(B) + 6h(A) + 3 \log 2$, hence the desired bounds. □

Note that since we assume that F generates a non virtually solvable group, then according to Theorem 1.2, $h(F) \geq e(F) \geq \varepsilon$ for some fixed ε. Therefore there exists a constant $D_{qs} > 0$ such that $\delta(F) \leq D_{qs}|F|^2 h(F)$.

Lemma 7.3. There is an integer $n_0 \geq 2$ such that if F is a finite subset of $SL_2(\mathbb{C})$ containing 1 and generating a non virtually solvable group, then for any semisimple $A \in F$ there exists $B \in F^{n_0}$ which is in nice position with respect to A.

Proof. This is another occurrence of the escape trick described in Lemma 4.8. The subvarieties $X_A = \{B \in GL_2,\ Bv_A^+ \in \{v_A^\pm\}$ or $Bv_A^- \in \{v_A^\pm\}\}$ are conjugate to each other in GL_2. In particular there is N as in Lemma 4.8 such that for each semisimple A in F, F^N is not contained in $X_A(\mathbb{C})$, as the group generated by F clearly cannot be contained in any $X_A(\mathbb{C})$ for it would otherwise be virtually solvable. □

We have for all $n \in \mathbb{N}$

$$\delta(F^n) \leq D_{qs} \cdot |F^n|^2 \cdot h(F^n) \leq D_{qs} \cdot |F|^{2n} \cdot n \cdot h(F)$$

We may write with obvious notation

$$\delta = \sum_{p\in\{\infty\}\cup\mathcal{P}} \delta_p = \delta_\infty + \delta_f$$

We fix $n = n_0$ as in Lemma 7.3 and let $D'_{qs} = D_{qs} \cdot n_0$ so that $\delta(F^{n_0}) \leq D'_{qs} \cdot |F|^{2n_0} \cdot h(F)$ and $h(F) \leq C_{qs} \cdot e(F)$. For each $p \in \{\infty\} \cup \mathcal{P}$ we set $e_p = e_p(F)$, $h_p = h_p(F)$ and $\delta_p = \delta_p(F^{n_0})$. We now claim:

Claim : *There exists a constant $C'' > 0$ such that for any set F in $SL_2(\overline{\mathbb{Q}})$ containing* 1 and *generating a non virtually solvable subgroup, there exist $p \in \{\infty\}\cup\mathcal{P}$ and a place $v|p$ such that, $\max\{\delta_v, h_v\} \leq C'' \cdot |F|^{n_0} \cdot e_v$ and $e_v > \frac{e_p}{2}$. Moreover if $p = \infty$, we may assume that $e_\infty \geq \frac{1}{2}e$.*

We now prove this claim. Suppose first that $e_\infty \geq \frac{1}{2}e$, then $\delta_\infty + h_\infty \leq C_{qs}(D'_{qs}|F|^{2n_0} + 1) \cdot e_\infty$. But

$$e_\infty \leq \frac{2}{[K:\mathbb{Q}]} \sum_{v\in V_\infty^+} n_v e_v$$

where $V_\infty^+ = \{v \in V_\infty,\, e_v \geq \frac{e_\infty}{2}\}$. Indeed

$$e_\infty = \frac{1}{[K:\mathbb{Q}]}\left(\sum_{v\in V_\infty^+} n_v e_v + \sum_{v\in V_\infty^-} n_v e_v\right) \leq \frac{1}{[K:\mathbb{Q}]}\sum_{v\in V_\infty^+} n_v e_v + \frac{e_\infty}{2}.$$

Hence $\sum_{v\in V_\infty^+} n_v(\delta_v + h_v) \leq 4C_{qs}(D'_{qs}|F|^{2n_0} + 1) \cdot \sum_{v\in V_\infty^+} n_v e_v$. So for at least one $v \in V_\infty^+$ we have $\max\{\delta_v, h_v\} \leq \delta_v + h_v \leq 4C_{qs}(D'_{qs}|F|^{2n_0} + 1) \cdot e_v$.

Now suppose $e_\infty < \frac{e}{2}$, then $e_f \geq \frac{e}{2} > 0$ and $\sum_{p\in\mathcal{P}} \delta_p + h_p \leq 2C_{qs}(D'_{qs}|F|^{2n_0} + 1) \cdot \sum_{p\in\mathcal{P}} e_p$ hence there must be one $p \in \mathcal{P}$ for which $e_p > 0$ and $\delta_p + h_p \leq 2C_{qs}(D'_{qs}|F|^{2n_0} + 1) \cdot e_p$. As this is an average over the places $v|p$, as before there must be some place $v|p$ for which $e_v \geq \frac{e_p}{2}$ and $\max\{\delta_v, h_v\} \leq \delta_v + h_v \leq 4C_{qs}(D'_{qs}|F|^{2n_0} + 1) \cdot e_v$. So we have justified the claim.

End of the proof of Theorems 1.3 and 1.1: Let us recapitulate what we have so far. We started with a set F in $SL_2(\overline{\mathbb{Q}})$ containing 1 and generating a non virtually solvable subgroup. We found the constant $n_0 \geq 2$ as in Lemma 7.3. We also found a constant C'' such that for some prime p and a place $v|p$ such that $\max\{\delta_v(F^{n_0}), h_v(F)\} \leq C'' \cdot |F|^{2n_0} \cdot e_v(F)$, and $e_v(F) \geq \frac{1}{4}e_p(F) > 0$ (with $e_\infty \geq \frac{e}{2}$ in case $p = \infty$). Set $D''_F := C'' \cdot |F|^{2n_0}$.

Suppose first that $v \in V_f$. Recall that we had $\Lambda_v(F^2) \geq E_v(F)^2$ by Lemma 2.1. Let $A_0 \in F^2$ be such that $\Lambda_v(A_0) = \Lambda_v(F^2)$. Then $\Lambda_v(A_0) \geq E_v(F)^2 \geq ||F||_v^{\frac{2}{D''_F}} > 1$ and hence if $k_1 \in \mathbb{N}$ is the first even integer strictly larger that D''_F, we have

$\Lambda_v(A) > ||F||_v$ if $A = A_0^{k_1/2} \in F^{k_1}$. Moreover we have $\delta_v(F^{n_0}) \leq D''_F \cdot e_v(F)$, therefore for every $B \in F^{n_0}$ which is in nice position with respect to A_0 (and there are such B's according to Lemma 7.3) we have $\delta_v(Id; A_0) + \delta_v(B; A_0) \leq D''_F \cdot e_v(F)$. Fix one such B. We have $d_v(Bv_A^{\varepsilon}, v_A^{\varepsilon'}) \geq E_v(F)^{-D''_F} \geq ||F||_v^{-D''_F}$ and also $d_v(v_A^{\varepsilon}, v_A^{\varepsilon'}) \geq E_v(F)^{-D''_F} \geq ||F||_v^{-D''_F}$ for all $\varepsilon, \varepsilon' \in \{\pm\}$. Therefore we are in a position to apply the ping-pong lemma 6.1 to the pair A and B with k_1 as above $(\leq D''_F + 2)$, $k_2 = n_0$ and $k_3 = D''_F$.

Suppose now that $v \in V_\infty$. We have $E_v(F) \geq \exp(\frac{e}{2}) \geq \exp(\frac{\varepsilon}{2})$ where ε is the constant from Theorem 1.2. Now Lemma 2.6 shows that there is a constant $n_1 = n_1(\varepsilon) \in \mathbb{N}$ such that $E_v(F^{n_1}) \geq \frac{2}{c^2}$ where c is the constant in Lemma 2.1. Then by Lemma 2.1 $\Lambda_v(F^{2n_1}) \geq c^2 E_v(F^{n_1})^2 \geq 2E_v(F^{n_1}) \geq 2E_v(F) \geq 2||F||^{\frac{1}{D''_F}}$. Observe that after possibly changing n_0 we may assume that it is larger than $2n_1$. Pick $A_0 \in F^{2n_1}$ such that $\Lambda_v(A_0) = \Lambda_v(F^{2n_1})$. Finally if k'_1 is the smallest integer strictly larger than D''_F, we set $A = A_0^{k'_1} \in F^{k_1}$ where $k_1 = 2n_1 k'_1$. We have $\Lambda_v(A) > 2||F||_v$. Moreover $\delta_v(F^{n_0}) \leq D''_F \cdot e_v(F)$, therefore for every $B \in F^{n_0}$ which is in nice position with respect to A_0 (and there are such B's according to Lemma 7.3) we have $\delta_v(Id; A_0) + \delta_v(B; A_0) \leq D''_F \cdot e_v(F)$. Fix one such B. We have $d_v(Bv_A^{\varepsilon}, v_A^{\varepsilon'}) \geq E_v(F)^{-D''_F} \geq ||F||_v^{-D''_F}$ and also $d_v(v_A^{\varepsilon}, v_A^{\varepsilon'}) \geq E_v(F)^{-D''_F} \geq ||F||_v^{-D''_F}$ for all $\varepsilon, \varepsilon' \in \{\pm\}$. Therefore we are in a position to apply the ping-pong lemma 6.1 to the pair A and B with k_1 as above $(\leq 2n_1(D''_F + 1))$, $k_2 = n_0$ and $k_3 = D''_F$.

Theorem 1.3 on the quasi-isometric embedding of the free group (in the case $F \subset SL_2(\overline{\mathbb{Q}})$) now follows readily by application of Lemma 6.3. To complete the proof of Theorem 1.1, it remains to observe that we can reduce to the situation where F has three elements $\{1, a, b\}$ by application of Lemma 4.7. Note that we cannot do this reduction for Theorem 1.3 because we need there to control the behaviour of every element of F.

There are several ways to see that Theorems 1.3 and 1.1 for $SL_2(\overline{\mathbb{Q}})$ imply the same theorems for $SL_2(\mathbb{C})$. One can use the remark made in the introduction that both results are equivalent to a countable union of assertions expressible in first order logic. By elimination of quantifiers for algebraically closed fields, we know that two algebraically closed fields of the same characteristic satisfy the same statements of first order logic (see e.g. [14] chp. 9). Hence the validity of Theorems 1.3 and 1.1 over $\overline{\mathbb{Q}}$ is equivalent to its validity over $\mathbb{C}$.

Another way to see it is to invoke Proposition 9.3 below and use the fact that if V is an algebraic variety defined over $\mathbb{Q}$, then $V(\overline{\mathbb{Q}})$ is Zariski-dense in $V(\mathbb{C})$. From Proposition 9.3 and the above proof over $\overline{\mathbb{Q}}$, we know that $\mathcal{W}_n(\overline{\mathbb{Q}}) \subset \mathcal{V}(\overline{\mathbb{Q}})$ for every $n \in \mathbb{N}$, which readily implies that $\mathcal{W}_n(\mathbb{C}) \subset \mathcal{V}(\mathbb{C})$ for every $n \in \mathbb{N}$ ($\mathcal{W}_n$ is defined in (15), see Section 9 below). And the theorem is proved over $\mathbb{C}$ with the same constant N_0.

For Theorem 1.1, one could also use a specialization argument as is in [13] for instance.

8. Uniform spectral gap in ℓ^2 and co-growth of subgroups

We prove here Corollaries 1.4 to 1.7. Corollary 1.4 is a direct application of Corollary 1.5, so we will not say more about it.

Proof of Corollary 1.5. We reproduce the argument given in [29] and [9]. Since the free group F_2 is non amenable, there is a constant $\kappa > 0$ such that $\max\{||a \cdot f - f||_2, ||b \cdot f - f||_2\} \geq \kappa \cdot ||f||_2$ for every $f \in \ell^2(F_2)$ where a and b are the two free generators of F_2. Then according to Theorem 1.1, there are a and $b \in (F \cup F^{-1})^{N_0}$ such that a and b generate a free subgroup H. For $f \in \ell^2(\Gamma)$ and Hx a coset of H, let f_{Hx} denote the restriction of f to Hx. Let $\mathcal{A}$ (resp. $\mathcal{B}$) be the subset of $H\backslash\Gamma$ of those cosets such that $||a \cdot f_{Hx} - f_{Hx}||_2 \geq \kappa||f_{Hx}||_2$ (resp. $||b \cdot f_{Hx} - f_{Hx}||_2 \geq \kappa||f_{Hx}||_2$). And set $f_{\mathcal{A}} = \sum_{Hx\in\mathcal{A}} f_{Hx}$, and $f_{\mathcal{B}} = \sum_{Hx\in\mathcal{B}} f_{Hx}$. Since $||f||_2^2 \leq ||f_{\mathcal{A}}||_2^2 + ||f_{\mathcal{B}}||_2^2$ we may assume without loss of generality that $||f_{\mathcal{A}}||_2^2 \geq ||f||_2^2/2$. Hence $\|a \cdot f - f\|_2^2 \geq \|a \cdot f_{\mathcal{A}} - f_{\mathcal{A}}\|_2^2 \geq \kappa^2\|f_{\mathcal{A}}\|_2^2 \geq \frac{\kappa^2}{2}\|f\|_2^2$. Since $a \in (F \cup F^{-1})^{N_0}$ we have

$$\|a \cdot f - f\|_2 \leq \sum_{i=1}^{N_0} \|s_1 \dots s_i \cdot f - s_1 \dots s_{i-1} \cdot f\|_2 = \sum_{i=1}^{N_0} \|s_i \cdot f - f\|_2$$

where $a = s_1 \cdot ... \cdot s_{N_0}$ with $s_i \in F \cup F^{-1}$. Finally, for some i we have $\|s_i \cdot f - f\|_2 = \|s_i^{-1} \cdot f - f\|_2 \geq \frac{\kappa}{N_0\sqrt{2}}\|f\|_2$. Hence we have proved the first assertion of Corollary 1.5 with $\varepsilon = \frac{\kappa}{N_0\sqrt{2}}$.

To prove the second assertion, let F and A be as in the statement. Let Γ be the group generated by A and F and simply apply the above with f the indicator function of A in $\ell^2(\Gamma)$.

Proof of Corollary 1.6. Set $\Gamma = \langle F\rangle$ with $F = \{a_1^{\pm1}, ..., a_m^{\pm1}\}$ and as in the statement and $\mu = \frac{1}{2m}\sum_{1\leq i\leq m} \delta_{a_i} + \delta_{a_i^{-1}}$. Then $\mathbb{P}(S_n = e) = \mu^n(e)$. But $\mu^{nN_0}(e) \leq ||\mu^{\frac{nN_0}{2}}||^2 \leq ||\mu^{N_0}||^n$ where $||\cdot||$ is the norm of the convolution operator. Theorem 1.1 shows that $\mu^{N_0} = \alpha\mu_{F_2} + (1-\alpha)\nu$ for some probability measure ν, where $\mu_{F_2} = \frac{1}{4}(\delta_a + \delta_{a^{-1}} + \delta_b + \delta_{b^{-1}})$ and a, b are the free generators in F^{N_0}, and $\alpha = \frac{1}{(2m)^{N_0}}$. It follows that $||\mu^{N_0}|| \leq 1 - \alpha\tau$ if $||\mu_{F_2}|| = 1 - \tau < 1$. Hence the result.

Proof of Corollary 1.7. We keep the notation of the proof of Corollary decay. One can go from spectral gap to co-growth in a one to one fashion, thanks to the following formula (see [17], [10], [27])

$$(2m-1)^\eta + (2m-1)^{1-\eta} = (2m)^\theta, \tag{14}$$

where $\eta = \lim_{n \text{ even}} \frac{1}{n}\log_{2m-1}|W'_n|$ and $\theta = \lim_{n \text{ even}} \frac{1}{n}\log_{2m}|W_n|$, with W_n the set of paths of length n in the free group F_{2m} going from the identity to itself and W'_n is the set of elements in F_{2m} of length n that kill $(a_1, ..., a_m)$. Since $|W'_{p+q+2}| \geq$

$|W_p'||W_q'|$, we must have $|W_n'| \leq (2m-1)^{\eta(n+2)}$ for all $n \geq 1$. On the other hand $\mu^n(e) = \frac{|W_n|}{(2m)^n}$ and hence $(2m)^\theta \leq 2m||\mu^{N_0}||^{\frac{1}{N_0}} \leq 2m(1-\alpha\tau)^{\frac{1}{N_0}}$ if $||\mu_{F_2}|| = 1-\tau$. Hence $(2m)^\theta \leq 2m(1-\frac{\tau}{(2m)^{N_0}})^{\frac{1}{N_0}}$. Solving equation (14), we obtain $(2m-1)^\eta \leq 2m-1-\frac{\tau}{(2m)^{2N_0}}$. Hence $|W_n'| \leq (2m-1-\frac{\tau}{2(2m)^{2N_0}})^n$ for $n \geq n(m)$. We are done.

To see the converse, note that by Proposition 9.2 $F = \{a_1, ..., a_m\} \subset GL_d(\mathbb{C})$ generates a virtually solvable subgroup if and only if it contains a subgroup of index $< M$ which can be conjugated in the upper triangular matrices, hence is of solvable length $\leq d$. In particular $\Gamma = \langle F \rangle$ is a quotient of the free object on m generators, which we denote by S, in this variety of groups. S is virtually solvable, hence amenable and by the Kesten's criterion and (14) must satisfy $|W_n'| \geq (2m-1-\varepsilon)^n$ for every $\varepsilon > 0$ and all $n \geq n(\varepsilon)$. Since every relation in S is also a relation in Γ we are done.

9. Large Girth

Here we prove Corollaries 1.8 to 1.11. Let $\overline{K}$ be an algebraically closed field, $\mathcal{F}_d$ the flag variety in $\overline{K}^d$ ($\mathcal{F}_d = \mathbb{P}^1(\overline{K})$ if $d=2$) and let $\mathcal{V}_k$ be the set of k-tuples $(A_1, ..., A_k) \in GL_d(\overline{K})^k$ such that $\mathbf{A} = (A_1, ..., A_k)$ leaves invariant some subset $\{u_1, ..., u_M\}$ of M not necessarily distinct points of $\mathcal{F}_d$.

Proposition 9.1. Then $\mathcal{V}_k$ is a closed subscheme of $GL_d(\overline{K})^k$ defined over $\mathbb{Z}$.

Proof. We write the proof for $k=2$. Consider the map $\phi : GL_d \times GL_d \times \mathcal{F}_d^M \to \mathcal{F}_d^{3M}$ which maps $(A, B, u_1, ..., u_M)$ to $(Au_1, ..., Au_M, Bu_1, ..., Bu_M, u_1, .., u_M)$. For every permutations $\sigma, \eta \in \mathcal{S}_M$ we set $\Delta_{\sigma,\eta} = \{(a_1, ..., a_M, b_1, ..., b_M, u_1, ..., u_M) \in \mathcal{F}_d^{3M}$ such that $a_i = u_{\sigma(i)}$ and $b_i = u_{\eta(i)}$ for each $i = 1, ..., M\}$ and let Δ the union of all $\Delta_{\sigma,\eta}$. Then Δ is a closed subvariety of $\mathcal{F}_d^{3M}$, therefore so is $\mathcal{V}_k = \pi \circ \phi^{-1}(\Delta)$, where π is the projection onto the $GL_d \times GL_d$ factor, which is a closed morphism since $\mathcal{F}_d^{3M}$ is complete. □

Proposition 9.2. (Zariski closedness of virtually solvable tuples) There is $M = M(d) \in \mathbb{N}$ such that a k-tuple $\mathbf{A} = (A_1, ..., A_k)$ in $GL_d(\mathbb{C})$ generates a virtually solvable subgroup if and only if $F \in \mathcal{V}_k(\mathbb{C})$.

Proof. The if part is clear. To show the converse observe that by induction on d we may assume that $\mathbb{G}$ acts irreducibly on $\mathbb{C}^d$. Since the connected component $\mathbb{G}_0$ is solvable, Borel's fixed point theorem implies that it fixes a point on $\mathcal{F}_d$. Let $\mathbb{U}$ be the unipotent radical of $\mathbb{G}_0$. If $\mathbb{U}$ is non trivial it must fix pointwise a non trivial subspace of $\mathbb{C}^d$. As $\mathbb{G}$ normalizes $\mathbb{U}$, $\mathbb{G}$ also must fix that subspace, which contradicts the assumption of irreducibility. Hence $\mathbb{U}$ is trivial and $\mathbb{G}_0$ is a torus. Therefore $\mathbb{G}$ is contained in the normalizer $N(\mathbb{G}_0)$ and $N(\mathbb{G}_0)/Z(\mathbb{G}_0)$ embeds in the Weyl group of GL_d hence has size at most $d!$. We may thus assume that $\mathbb{G}$ centralizes $\mathbb{G}_0$. As we may again assume that $\mathbb{G}$ acts irreducibly, this forces $\mathbb{G}_0$ to be trivial. Hence we are left with the case when $\mathbb{G}$ is finite and we invoke Jordan's theorem (see e.g. [11]) to conclude: it gives $M \in \mathbb{N}$ such that $[\mathbb{G} : A] \leq M$ where A is an abelian subgroup of $GL_d(\mathbb{C})$ made of semisimple elements. Hence A is contained in a torus S, which fixes a flag. It follows that $\mathbb{G}$ stabilizes the $\mathbb{G}$-orbit of this flag, which is of cardinality at most M. □

Let us now express the conclusion of Theorem 1.1 in terms of a countable family of algebraic conditions. Let N be the integer obtained in the statement of Theorem 1.1 and let $B_2(n)$ be the ball of radius n in the free group F_2 on two generators. For $n \geq 1$ let $\mathcal{W}_n$ be the set of k-tuples $\mathbf{A} = (A_1, ..., A_k) \in GL_d(\mathbb{C})^k$ such that for any words w_1 and w_2 in $B_k(N)$ there exists a word $w \in B_2(n)\backslash\{1\}$ such that $w(w_1(\mathbf{A}), w_2(\mathbf{A})) = 1$. Clearly $\mathcal{W}_n$ is a closed subvariety of $GL_d(\mathbb{C})^k$. Hence we obtain:

Proposition 9.3. (reformulation of main theorem in terms of equality of algebraic varieties) Theorem 1.1 is equivalent to the statement: $\forall n \geq 1$ $\mathcal{W}_n \subset \mathcal{V}_k$.

Remark 9.4. Clearly $\mathcal{W}_n \subset \mathcal{W}_{n+1}$. Also it is clear from Proposition 9.2 that $\mathcal{V}_k \subset \mathcal{W}_{n_0}$ for some $n_0 \geq 1$. Hence Theorem 1.1 is in fact equivalent to $\mathcal{W}_n = \mathcal{W}_{n_0} = \mathcal{V}_k$ for all $n \geq n_0$.

For $w \in F_k\backslash\{1\}$ let X_w be the word variety $X_w = \{\mathbf{A} \in GL_d(\mathbb{C})^k,\ w(\mathbf{A}) = 1\}$. Equivalently $X_w = \{\mathbf{A} \in GL_d(\mathbb{C})^k,\ Q_w^{ij} = 0 \text{ for all } i, j = 1, ..., d\}$, where $Q_w^{ij} = P_w^{ij} - \delta_{ij}$ and $\{P_w^{ij}\}_{1 \leq i,j \leq d}$ is the matrix $w(\mathbf{A})$ with each $P_w^{ij} \in \mathbb{C}[((A_1)_{ij})_{ij}, ..., ((A_k)_{kl})_{kl}]$ a polynomial in the kd^2 variables of $\mathbf{A} = (A_1, ..., A_k)$. Let $\mathcal{A}$ be the set of couples (w_1, w_2) of words in $B_2(N)$. Let $\mathcal{B}_n$ be the set of words $w \in B_k(n)\backslash\{1\}$ and finally let $\mathcal{C}$ be the set of indices $\{ij\}_{1 \leq i,j \leq d}$. For $a = (w_1, w_2) \in \mathcal{A}$, $b = w \in \mathcal{B}_n$ and $c = \{ij\} \in \mathcal{C}$ set $Q_{a,b,c}$ to be the polynomial $Q^{ij}_{w(w_1,w_2)}$.

Lemma 9.5. (degree and height bounds for word polynomials) For each $a \in \mathcal{A}$, $b \in \mathcal{B}_n$, and $c \in \mathcal{C}$, the polynomial $Q_{a,b,c} \in \mathbb{Z}[((A_1)_{ij})_{ij}, ..., ((A_k)_{kl})_{kl}]$ has integer coefficients, has height at most $d^{nN} + 1$ and degree at most nN.

Proof. Here the height is understood in the naive sense of maximal modulus of the coefficients. The proof is an easy induction on n and we omit the details. □

With this notation, we have $\mathcal{W}_n = \cap_{a \in \mathcal{A}} \cup_{b \in \mathcal{B}_n} \cap_{c \in \mathcal{C}} \{Q_{a,b,c} = 0\}$, which we may rewrite as

$$\mathcal{W}_n = \bigcup_{f \in \mathcal{B}_n^{\mathcal{A}}} \mathcal{W}_{n,f} \tag{15}$$

where $\mathcal{W}_{n,f} = \cap_{a \in \mathcal{A}} \cap_{c \in \mathcal{C}} \{Q_{a,f(a),c} = 0\}$ where f ranges among all maps $f : \mathcal{A} \to \mathcal{B}_n$. Let I be the ideal of $\mathbb{Z}[(A_{ij})_{ij}, (B_{kl})_{kl}]$ associated to $\mathcal{V}$. Let $I_{n,f}$ be the ideal of $\mathbb{Q}[(A_{ij})_{ij}, (B_{kl})_{kl}]$ generated by the $Q_{a,f(a),c}$ with $a \in \mathcal{A}$ and $c \in \mathcal{C}$. Let I_n^f be the ideal of all polynomials in $\mathbb{Q}[(A_{ij})_{ij},(B_{kl})_{kl}]$ that vanish on $\mathcal{W}_{n,f}$. Then Hilbert's Nullstellensatz asserts that $I_n^f = \sqrt{I_{n,f}}$, and Theorem 1.1 says that $I \subset I_n^f$ for every n and $f \in \mathcal{B}_n^{\mathcal{A}}$. Let $f_1,...,f_m$ be generators of I with integer coefficients. The following effective version of the Nullstellensatz may be found in [26]:

Theorem 9.6. (**Effective arithmetic nullstellensatz** [26]) Let $r,d \in \mathbb{N}$, $h>0$ and $f,q_1,...,q_k$ be polynomials in $\mathbb{Z}[X_1,...,X_r]$ with logarithmic height at most h and degree at most d. Assume that f vanishes at all common zeros (if any) of $q_1,...,q_k$ in $\mathbb{C}[X_1,...,X_r]$. Then there exist $a,e \in \mathbb{N}$ and polynomials $b_1,...,b_k \in \mathbb{Z}[X_1,...,X_r]$ such that

$$af^e = b_1q_1 + ... + b_kq_k$$

with $e \leq (8d)^{2^r}$, the total degree of each b_i at most $(8d)^{2^r+1}$ and the logarithmic height of each b_i as well as a is at most $(8d)^{2^{r+1}+1}(h+8d\log(8d))$.

Here, the logarithmic height is the log of the naive height used above. We can now finish the proof of Corollary 1.8. In our situation, Theorem 9.6 yields numbers $a_i \in \mathbb{N}$ and polynomials $b^i_{f,a,c} \in \mathbb{Z}[(A_{ij})_{ij},(B_{kl})_{kl}]$ with logarithmic height $h_n = O_d(n^{2^{kd^2+2}})$ as well as numbers $e_i \in \mathbb{N}$ such that for each $i=1,...,m$

$$a_i f_i^{e_i} = \sum_{a\in\mathcal{A},c\in\mathcal{C}} b^i_{f,a,c} Q_{a,f(a),c} \tag{16}$$

It follows that if $p > \exp(h_n)$ is a rational prime, then for any field K of characteristic p, and any $\mathbf{A} \in GL_d(K)^k$, if for any words w_1,w_2 in $B_k(N)$ there is a word $w \in B_2(n)\backslash\{1\}$ such that $w(w_1(\mathbf{A}),w_2(\mathbf{A}))=1$, then $f_i(\mathbf{A})=0$ for all $i=1,...,m$. Since the f_i generate I, according to Proposition 9.1, this means that there must be a set $\{u_1,...,u_M\}$ in $\mathcal{F}_d(\overline{K})$ of at most M points ($\overline{K}$ is an algebraic closure of K) which is fixed by $\mathbf{A}$. In particular the group Γ generated by $\mathbf{A}$ contains the solvable subgroup $\Gamma_0 = \{\gamma \in \Gamma, \gamma\cdot u_1 = u_1\}$ as a subgroup of index $\leq M$. Therefore Corollary 1.8 holds as soon as $p > \exp(h_n)$, i.e. for all $n \leq O_d(\log p)^{2^{-kd^2-2}}$. This ends the proof of Corollary 1.8 with e.g. $\varepsilon_0 = 2^{-10}$ for $k=d=2$.

Observe that in the above proof we may have replaced $\mathcal{W}_n$ by the larger subvariety $\mathcal{W}_n'$ equal to the subset of k-tuples $\mathbf{A}=(A_1,...,A_k) \in GL_d(\mathbb{C})^k$ such that for any words w_1 and w_2 in $B_k(N)$ there exists a word $w \in B_2(n)\backslash\{1\}$ such that $w(w_1(\mathbf{A}),w_2(\mathbf{A}))=1$ or there are words $w \in B_2(n)\backslash\{1\}$ and $w_0 \in B_k(\frac{1}{4d}\ell(w))$ such that $w(w_1(\mathbf{A}),w_2(\mathbf{A}))=w_0(\mathbf{A})$. The bounds on the height and degree of the polynomials defining $\mathcal{W}_n'$ are of the same magnitude as those of $\mathcal{W}_n$, hence the same conclusion holds, namely:

Corollary 9.7. (**QI embedded subgroup of large girth**) Given $d,k \geq 2$, there is $N,M \in \mathbb{N}$ and $\varepsilon_0, C > 0$ such that for every prime p and every field K of characteristic p and any finite subset F with k elements generating a subgroup

G of $GL_d(K)$ which contains no solvable subgroup of index at most M, then F^N contains two elements a, b such that $w(a,b) \neq 1$ in $GL_d(K)$ generating a subgroup H with Cayley graph $\mathcal{G}_H$ such that $girth(\mathcal{G}_H) \geq f(p)$ and

$$\frac{1}{C} \cdot d_G(1,h) \leq d_H(1,h) \leq C \cdot d_G(1,h) \tag{17}$$

for any h with $d_H(1,h) \leq f(p)$, where $f(p) = (\log p)^{\varepsilon_0}$. Note that C depends on d only.

Proof of Corollary 1.9. This follows directly from Corollary 9.7 and the following Proposition:

Proposition 9.8. (QI embedded subgroup of large girth implies uniform expansion on small sets) Suppose G is a k-generated group with Cayley graph $\mathcal{G}_G$ and word metric d_G and H is a finitely generated subgroup with Cayley graph $\mathcal{G}_H$ and word metric d_H. Assume further that $girth(\mathcal{G}_H) \geq N$ and that (17) holds for all h such that $d(1,h) \leq N$. Let μ be the uniform symmetric probability measure on the generators of G. Then there exist an explicit constant $\beta = \beta(k, C) > 0$ such that

$$||\mu * f||_2 \leq (1-\beta)||f||_2$$

for any function f supported on a ball of radius $\leq N/2C$ in $\mathcal{G}_G$.

Note that the uniform QI-embedding of H in G is used in a key way in the proof. This Proposition also yields in a standard way the following corollary.

Corollary 9.9. In the setting of the Proposition, there is $\beta = \beta(C,k) > 0$ and $\alpha = \alpha(C,k) > 0$ such that $\mu^{*n}(e) \leq (1-\beta)^n$ for all $n \leq N/2C$. Moreover for any subset A of G lying in a ball of radius $\leq N/2C$, there is a generator s such that $|sA \triangle A| \geq \alpha|A|$.

Proof of Proposition 9.8. Let $B_G(N)$ be the ball of radius N in G centered at 1. Pick representatives $rep = \{\overline{x} \in B_G(N)\}$ for right cosets of H : $H \cdot B_G(N) = \cup_{\overline{x} \in rep} H\overline{x}$ and then split f as a sum of mutually orthogonal terms

$$f = \sum_{x \in rep} f_{\overline{x}}(\cdot \overline{x}^{-1})$$

where $f_{\overline{x}} : H \to \mathbb{R}$ send h to $f(h\overline{x})$. We have $||f||^2_{2,G} = \sum_{rep} ||f_{\overline{x}}||^2_{2,H}$. Let $S = B_H(1)$ the generating set for H. We know from the corresponding spectral gap estimate on the free group that for every $g : H \to \mathbb{R}$ such that $Supp(g) \subset B_H(N)$ there exists $s \in S$ such that $||s \cdot g - g||_{2,H} \geq \tau ||g||_{2,H}$ where $\tau > 0$ is an absolute constant (independent of the rank of the free group). But if $f_{\overline{x}}(h) \neq 0$ then $h\overline{x} \in Supp(f) \subset B_G(N/2C)$, hence by (17), $h \in B_H(N)$, so $Supp(f_{\overline{x}}) \subset B_H(N)$. Hence there is $s_{\overline{x}} \in S$ such that $||s_{\overline{x}} \cdot f_{\overline{x}} - f_{\overline{x}}||_{2,H} \geq \tau ||f_{\overline{x}}||_{2,H}$. We get:

$$\sum_{s \in S} ||s \cdot f - f||^2_{2,G} = \sum_{\overline{x} \in rep} \sum_{s \in S} ||s \cdot f_{\overline{x}} - f_{\overline{x}}||_{2,H} \geq \tau^2 \cdot ||f||^2_{2,G}$$

Now since $d_G(1, s) \leq C$ for each $s \in S$ by (17), we get:

$$||s \cdot f - f||_{2,G}^2 \geq \frac{\tau^2}{|S|}||f||_{2,G}^2$$

note that $|S| \leq |B_G(C)| \leq (2k)^C$. From there it is straightforward to derive the bound:

$$||\mu * f||_2 \leq \frac{1}{2}\left[1 + \sqrt{\left(1 - \frac{\tau^2}{32(2k)^{2C}}\right)}\right] \cdot ||f||_2.$$

Proof of Corollary 1.10. Applying Lemma 4.7, we may assume that $Card(F) = 2$. If $F = \{A, B\}$ does not satisfy the conclusion of the corollary for N as in Theorem 1.1, then for arbitrarily small $\delta > 0$ there is a map $f : \mathcal{A} \to \mathcal{B}_{n(\delta)}$ where $n = n(\delta) \leq (\log \delta^{-1})^{\varepsilon_1}$ (with ε_1 an absolute constant to be determined below) such that $d(f(a), 1) < \delta$ for all $a \in \mathcal{A}$. This means that for some $C_1 > 0$ (depending on the choice of the Riemannian metric d), we have $\forall a \in \mathcal{A}\ \forall c \in \mathcal{C}$, $|Q_{a,f(a),c}| < C_1 \cdot \delta$. Applying Theorem 9.6 we get (16) as above. Moreover the logarithmic height of the polynomials $b^i_{f,a,c}$ is at most $O_d(n^{2^{2d^2+2}})$, hence evaluated on $F = \{A, B\}$, $b^i_{f,a,c}$ is $O_{d,F}(n^{2^{2d^2+2}})$, and hence also $f_i^{e_i} = O_{F,d}(\delta \cdot \exp(n^{2^{2d^2+2}}))$. As $e_i = O(n^{2^{2d^2}})$, we see f_i is arbitrarily small when $\delta \to 0$ as soon as we take for instance $\varepsilon_1 = 2^{-(2d^2+3)}$. It follows that $(A, B) \in \mathcal{V}_2$. A contradiction.

Proof of Corollary 1.11. With the notation of Corollary 1.7, define $\tau(m) > 0$ by $\frac{(2m-1-\frac{\varepsilon}{mD})}{(2m-1)} = e^{-\tau(m)}$ and let $\tau = \tau(2)$. By contradiction, suppose Corollary 1.11 does not hold for $(a, b) \in GL_d(\mathbb{C})$. Then for arbitrarily small δ, there is $n = n(\delta) \leq (\log \delta^{-1})^{\varepsilon_1}$ for which one can find at least $3^n e^{-\tau n}$ reduced words w of length n such that $d(w, 1) < \delta$. Let $\mathcal{I}$ be the set of all subsets of cardinality $3^n e^{-\tau n}$ of reduced words of length n. For $I \in \mathcal{I}$, let $\mathcal{V}_I$ the subvariety of $GL_d(\mathbb{C})^2$ where all words in I vanish simultaneously. Then Corollary 1.7 can be reformulated as the statement $\mathcal{V}_I \subset \mathcal{V}$ for every $n \geq n(2)$ and every $I \in \mathcal{I}$. As above the effective Nullstellensatz applies and gives coefficients $a_i \in \mathbb{Z}\backslash\{0\}$, polynomials $b^k_{I,w,ij}$ with coefficients in $\mathbb{Z}$, of degree and logarithmic height $O(n^{2^{2d^2+2}})$ and integers $e_i = O(n^{2^{2d^2}})$ such that

$$a_k f_k^{e_k} = \sum_{w \in I, \{ij\} \in \mathcal{C}} b^k_{I,w,ij} Q^{ij}_w$$

It follows that $f_k^{e_k} = O(\delta \cdot \exp(n^{2^{2d^2+2}}))$, which again implies that f_k is arbitrarily small as $\delta \to 0$ if $\varepsilon_1 = 2^{-2d^2-3}$, say. Hence $(a, b) \in \mathcal{V}_2$, a contradiction. Q.E.D.

Acknowledgments 9.10. I am very grateful to P. Sarnak for his precious tips and his enthusiasm. I thank E. Bombieri, J-B. Bost, A. Chambert-Loir and A. Yafaev for sharing their insights about diophantine geometry. I am grateful to

J. Bourgain and A. Gamburd for their encouragement and for pointing out the right girth bound in the corollary about $SL_2(\mathbb{F}_p)$. I thank P.E. Caprace, G. Chenevier, E. Lindenstrauss, G. Prasad and A. Salehi-Golsefidy for our stimulating conversations. Last but not least, I want to express my sincere and deep gratitude to T. Gelander for the time we spent doing mathematics together, which greatly contributed to my involvement in the questions addressed in this paper.

References

[1] Alperin R. and Noskov G., *Uniform Exponential Growth, Actions on Trees and GL_2*, AMS Contemporary Math. **298** (2002).

[2] Bilu, Y, *Limit distribution of small points on algebraic tori*, Duke Math. J. **89** (1997), no. 3, 465–476.

[3] Bombieri, E., and Gubler, W., *Heights in Diophantine geometry*, New Mathematical Monographs, **4**. Cambridge University Press, Cambridge, (2006).

[4] Bourgain, J. Gamburd, A., *Uniform expansion bounds for Cayley graphs of $SL_2(\mathbb{F}_p)$*, to appear in Annals of Math.

[5] Bourgain, J. Gamburd, A., *On the spectral gap for finitely generated subgroups of $SU(2)$*, to appear in Invent. Math.

[6] Breuillard, E., *On uniform exponential growth for solvable groups,* Pure and Applied Math. Quart. 3, no 4, Margulis Volume Part 1, 949–967, (2007).

[7] Breuillard, E, *An height gap theorem for the finite subsets of $SL_n(\overline{\mathbb{Q}})$ and non amenable linear groups*, preprint April 2008, arXiv:0804.1391.

[8] Breuillard, E, *A Strong Tits alternative*, preprint April 2008, arXiv:0804.1395.

[9] Breuillard, E, and Gelander, T., *Uniform independence in linear groups*, Invent. Math. **173**, no 2, 225–263, (2008).

[10] Cohen, J.M., *Cogrowth and amenability of Discrete Groups*, J. Funct. Anal. **48** (1982), 301-309.

[11] C.W. Curtis, I. Reiner, *Representation Theory of Finite Groups and Associative Algebras*, (Interscience, New York) (1962).

[12] P. Erdös and P. Turan, *On the distribution of roots of polynomials*, Ann. of Math. (2) **51** (1950), 105–119.

[13] Eskin, A., Mozes, S, Oh, Hee, *On uniform exponential growth for linear groups*, Invent. Math. **160** (2005), no. 1, 1–30

[14] Fried, M. and Jarden, M., *Field arithmetic*, 2^{nd} edition, Ergebnisse der Math. Grenzg., **11**, Springer-Verlag, (2005).

[15] Gamburd, A., Jakobson, D., Sarnak, P., *Spectra of elements in the group ring of $SU(2)$*, J. Eur. Math. Soc. (JEMS) 1 (1999), no. 1, 51–85.

[16] Gamburd, A., Hoory S., Shahshahani M., Shalev A., Virag, B., *On the girth of random Cayley graphs,* arXiv preprint (2005).

[17] Grigorchuk, R, *Symmetrical Random Walks on Discrete Groups*, in Multicomponent Random Systems ed. Dobrushin, Ya. Sinai, Adv. Prob. Related Topics **6**, Dekker (1980), 285-325.

[18] R. Grigorchuk, P. de la Harpe, *Limit behaviour of exponential growth rates for finitely generated group*s, in Essays on geometry and related topics, Vol. 1, 2, 351–370, Monogr. Enseign. Math., **38**, (2001).

[19] Kaloshin, V. Rodnianski, I, *Diophantine properties of elements of $SO(3)$*, Geom. Funct. Anal. 11 (2001), no. 5, 953–970.

[20] Kazhdan, D., and Margulis, G., *A proof of Selberg's hypothesis*, Mat. Sb. (N.S.) **75**, 117 (1968) 163–168.

[21] Kesten, H., *Symmetric walks on groups*, Trans. Amer. Math. Soc. **92** (1959) 336–354.

[22] Lang, S., *Fundamentals of Diophantine geometry*, Springer-Verlag, New York, (1983).

[23] Lubotzky, A., *Discrete groups, expanding graphs and invariant measures,*. Progress in Math., Birkhauser 1994.

[24] Lubotzky A., Phillips R., Sarnak P., *Hecke operators and distributing points on the sphere I*, CPAM, **39** (1987), 149-186.

[25] M. Mignotte, *Sur un théorème de M. Langevin*, Acta Arith. **54** (1989), 81–86

[26] Masser, and Wustholz, *Fields of large transcendence degree generated by values of elliptic functions*, Invent. Math. 1983

[27] Ollivier, Y., *Cogrowth and spectral gap of generic groups*, Annales de l'institut Fourier, **55** no. 1 (2005), p. 289-317

[28] Sarnak, P. *Applications of modular forms*, Cambridge University Press.

[29] Y. Shalom, *Explicit Kazhdan constants for representations of semisimple and arithmetic groups*, Ann. Inst. Fourier, **50** (2000), no. 3, 833–863.

[30] Thurston, W, *Three-dimensional geometry and topology*, Vol. 1. Edited by Silvio Levy, Princeton Mathematical Series, **35**. Princeton University Press, (1997).

[31] J. Tits, *Free subgroups of Linear groups*, Journal of Algebra **20** (1972), 250-270.

[32] Zhang, S., *Small points and adelic metrics*, J. Algebraic Geom. 4 (1995), no. **2**, 281–300

Emmanuel Breuillard, Ecole Polytechnique, France
E-mail address: `emmanuel.breuillard@math.polytechnique.fr`